Author:
Hull, Ellinor C M

Title:
Semantic Cognition and Cognitive Control in School-Aged Children

Semantic Cognition and Cognitive Control in School-Aged Children

Author Name: Ellinor Hull

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master of Science by Research in the Faculty of Life Sciences, School of Psychological Science. Submitted: February 2021.

Word Count: (21094)

ABSTRACT

According to the Controlled Semantic Cognition (CSC) framework, meaningful knowledge needs to be represented and controlled in order for it to be used in everyday tasks (Lambon Ralph, Jefferies, Patterson & Rogers, 2016). Some populations, such as older adults and patients with Semantic Aphasia, have general difficulties with cognitive control, which can impact their ability to retrieve meaningful concepts (Hoffman, 2018; Jefferies & Lambon Ralph, 2006). Children are also typically characterised as having deficits with cognitive control and even though studies have probed how cognitive control affects conceptual processing in pre-schoolers, research has failed to address the CSC framework in developing populations directly. This thesis aims to explore semantic cognition in school-aged children using a computerised taxonomic 2 alternative forced choice semantic picture matching task (SPM) programmed with a mouse-tracking software. The SPM manipulated the semantic relations of picture-probes and picture-alternatives in four different trial-types to adjust the level of semantic control required to select the most related picture-target. The first hypothesis predicts that SPM trial-types that require more control will be less accurate, yield slower response times and produce larger geometric mouse-tracking scores compared to trial-types that require less control. The second hypothesis predicts that chronological age and vocabulary knowledge will help predict performance on the SPM, and the third hypothesis predicts that non-semantic cognitive control will help predict performance on the SPM as well. Overall, the findings revealed that the control demands manipulated in the SPM did affect accuracy, response duration and geometric mouse-tracking scores, and that chronological age, vocabulary knowledge and non-semantic inhibitory control could predict response duration performance in the SPM. The findings are discussed in terms of how developing populations can fit into the CSC framework and solutions for future studies are proposed.

Acknowledgements

I have received a great deal of support and assistance throughout the project, from setting up the experiment to writing up the thesis. I would first like to thank my supervisor, Dr Josie Briscoe, for all her time and vital input. I am grateful for your guidance, patience and expertise. Your involvement made this thesis possible.

I would also like to give a huge thank you to all the families and schools that took part in the project. The time you took to go out of your way from your daily schedule was invaluable and your enthusiasm for research reminds me that there are many phenomena, however big or small, that people want answers to.

Thank you to the Bristol Cognitive Development Centre and to the undergraduate students who helped with recruitment and testing.

Last but not least, thank you to my family, friends and colleagues who encouraged me every day to keep writing and stay positive in the last 10 months. It would have been difficult without you.

Covid-19 Statement

Covid-19 has mainly impacted participant recruitment and the contents of the thesis. I was scheduled to test participants diagnosed with Autism Spectrum Disorder in a school in April 2020, however, due to the first coronavirus lockdown, it was cancelled. Having a sample with Autism Spectrum Disorder was originally going to be used to further understand and argue that cognitive control processes influence semantic cognition (due to part of the diagnostics of Autism Spectrum Disorders). No attempts were made to adapt to online testing since I was already at the end of my testing year.

In addition, 10 typically developing participants had been scheduled in to take part in the study at the Bristol Cognitive Development Centre between 23rd of March 2020 and 19th of April 2020. A last recruitment attempt in this period was going to be made to achieve a total of 100 participants with complete data (not counting any exclusions). Again, no attempts were made to adapt to online testing since I was already at the end of my testing year. The additional participants could have achieved a more even spread of the age-bands, which would have helped the age analyses.

Lastly, due to unanticipated difficulties of working from home, two parts of the project were excluded in the final write-up. Firstly, a data analysis with seven participants diagnosed with Autism Spectrum Disorders was excluded. It was partly excluded due to the number of participants in the group and because of the time consumption of additional literature reviews and complex analyses (case series analysis). Secondly, a data analysis of a semantic verbal fluency task was also excluded due to the same time-consuming processes (literature reviews and analyses). The absence of these components in the thesis further affected the contents and argument.

Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

List of Contents

List of Tables and Figures

1. Introduction

Concepts are mental representations formed in the brain from acquiring verbal and non-verbal knowledge about the world. For example, when one hears the word carrot, it is easy to visualise an orange vegetable even though a carrot is not visually present. Some concepts will share properties or events with each other. A carrot can be related to a rabbit because rabbits eat vegetables, but a carrot can equally be related to a basketball because they can both be orange. These semantic associations between different concepts create semantic networks (Foltz, 2001) and are stored as semantic memory in the brain (Balota & Coane, 2008). Semantic networks address the interconnectivity of concepts whilst semantic memory allows the connections to be preserved. Past research has identified semantic cognition to be a broader term for any cognitive processes that concerns concept knowledge (Chapman, Hasan, Schulz & Martin, 2020). Whilst past research efforts have been made to investigate semantic cognition in older adults and adults with brain dysfunctions, no research to date has addressed semantic cognition directly in developing populations. This thesis will focus on establishing evidence for processes of semantic cognition in typically-developing school-aged children and whether factors that influence adults' semantic cognition may also predict children's semantic cognition.

1.1 Semantic Cognition and Cognitive Control in Adults

Semantic cognition allows collected knowledge to be stored, manipulated and later utilised in situational contexts (Rogers & McClelland, 2006). Semantic cognition is specifically reliant on how individuals construct conceptual representations and how well individuals can regulate what knowledge to activate (Hoffman, McClelland & Lambon Ralph, 2018). For example, the nutritional information of a carrot becomes irrelevant when it needs to be peeled, but information about the types of appropriate peeling utensils is relevant and should be activated in order for the peeling to be successful. Neuropsychological research with two types of adult patients with impairments in semantic cognition have separated two distinct processes of semantic cognition and provides an explanation into how semantic cognition works (Chapman et al., 2020). One of the disorders is Semantic Dementia (SD) and the symptoms of SD manifests as a constant degradation of semantic representations (Delacourte, 2009). Concepts that patients with SD could once mentally visualise become unclear and sometimes morphed into other concepts. For example, a study using sketch drawings as an experimental measure found that patients with SD who were tasked to draw a duck from memory or replicate a picture of a duck with a 15-minute delay ended up giving the bird four legs (Bozeat et al., 2003). In this example, the concept duck had morphed into another concept (e.g., dog) where four limbs is a logical assumption. Here, the lack of a discrimination between two concepts shows that a loss of concept representation can impede conceptual knowledge and performance in an ordinary activity (drawing).

Research investigating the source of SD has successfully localised the disorder to damage in the Anterior Temporal Lobe (ATL; Mummary et al., 2000) and functional Magnetic Resonance Imaging (fMRI) has identified activation of the ATL when even neurotypical adults engage in semantic tasks (Teige, 2017; Teige et al., 2019). The involvement of the ATL in SD in particular has allowed researchers to establish a hub-and-spoke model that explains how mental representations are created. The hub-and-spoke model (Patterson, Nestor & Rogers, 2007) proposes that knowledge is collected through modality-specific interactions and is encoded in the corresponding brain area for each modality. For instance, sound is processed in the superior temporal gyrus (Howard et al., 2000), vision is processed in the occipital lobe (Johnson et al., 2015), and object function is processed in the left medial temporal lobe (Chen, Garcea & Mahon, 2016). The modality-specific brain areas are referred to as the spokes in the model and allow for collected knowledge to be managed and stored. Each spoke is connected to a central transmodal hub, which functions as a mediator of knowledge. The hub is identified as the ATL and when a mental representation is required, the ATL receives appropriate knowledge for the concept from each spoke and interlinks the information to create a mental representation for the concept. For example, the knowledge pertaining to the concept carrot would involve many featural aspects (e.g., colour, size and texture), but also that it grows in the ground, many animals consume it (e.g., horses, rabbits, and guinea pigs), and it is commonly used as a nose when building snowmen. The connectivity between the spokes and the hub allows for a full-bodied semantic account to be made for a concept, which is activated when the concept is needed for a task or activity. The hub-and-spoke model further incorporates that if the connectivity between the hub and spokes are weakened, then it would impair the semantic representations that could be constructed. Therefore, the damage to the ATL in patients diagnosed with SD is what weakens the ability to construct a full-bodied account of a concept. However, even though the ability to create semantic representations is an important process in semantic cognition, researchers have identified that controlled processes are equally as vital when recruiting concept knowledge through studying another patient group (Corbett, Jefferies, Ehsan & Lambon Ralph, 2009). These further findings paved the way for a more all-inclusive theoretical framework for semantic representations and cognition.

Semantic Aphasia (SA) is the second disorder that provides evidence for semantic cognition in adults (Jefferies & Lambon Ralph, 2006). Physically, patients with SA are presented with damage in the prefrontal cortex, temporoparietal region or both (Gerdner et al., 2012) and mentally, the damage translates into a deficit in cognitive control abilities specific to conceptual knowledge (Badre, Hoffman, Cooney & D'Esposito, 2009). Cognitive control, also known as executive function or executive control, is defined as the ability to use a range of complex mental processes to manipulate and guide thinking (Doebel, 2020) and is usually separated into three different mental processes; working memory, inhibition and flexibility (Miyake et al., 2000). To investigate how behaviourally different patients diagnosed with SD and SA are in recruiting concept knowledge, one research study contrasted seven patients with SA and eight patients with SD in a semantic battery of tasks including the comprehension of object uses, word meanings and picture meanings (Corbett et al., 2009). The same test-items were used in the different tasks and since patients with SD are hindered by their semantic representations, they performed in a stable manner for all the test-items across the tasks; if the patient could successfully represent appropriate conceptual knowledge for the test-item, they would perform accurately. In contrast, patients with SA performed more sporadically because in some tasks they could recruit the concept knowledge for the test-item but in other task settings they could not. The findings suggested that the deficit in patients with SA was not with semantic representations but with controlling what knowledge to bring forward and when. The contrasting findings from patients with SA and SD prompted a revision of the hub-and-spoke model and also found researchers suggesting that the semantic cognition network extends across a large brain area rather than being limited to just the ATL, especially since the ATL is intact in SA (Jefferies, 2013).

To incorporate the role of controlled processing in the hub-and-spoke model, an extension of the theory was made to form the Controlled Semantic Cognition (CSC) Framework (Lambon Ralph, Jefferies, Patterson & Rogers, 2016). The CSC framework emphasises that there are separate, but interacting, neural networks for semantic representation and controlled processing of concept knowledge. In the model, the controlled processes are referred to as semantic control, which is defined as the regulation of activating semantic knowledge pertinent to task-specific needs (Rogers, Patterson, Jefferies & Lambon Ralph, 2015). Whilst semantic representations are created as the hub-and-spoke model explains, there is a semantic control network that mediates what representations are brought forward due to task specific requirements. For example, if the task is to peel a carrot, the hub-and-spoke model will bring forward all knowledge regarding carrots and how to peel them, however, not all of the associated knowledge is relevant and therefore, according to the CSC framework, semantic control filters what semantic representations are appropriate for the completion of the task. Due to classifying the impaired controlled processes in patients with SA as semantic control, the CSC framework proposes that patients with SA will be unable to regulate what specific semantic representations to situationally bring forward despite having the knowledge to successfully represent the concept. Even though the CSC framework encompasses the semantic shortcomings of SA, researchers have started to probe more questions about semantic control and how it differs from general cognitive control abilities in other populations.

To test whether semantic control is dissimilar to cognitive control, Thompson and colleagues (2018) compared conceptual processing in 24 patients with SA and conceptual processing in 12 patients with consistent cognitive control disabilities (not SA). All patients completed a battery of semantic assessments that manipulated control demands, called the Cambridge Semantic Battery (Bozeat et al., 2000), and the two patient-groups were contrasted. The study identified that the inability to regulate semantic concepts was not unique to SA, but that general cognitive control abilities also play a role in conceptual processing. Patients with SA were largely more impaired on the semantic assessments compared to the other patient group, but both groups showed similar relationships in performance. In the battery, both groups were unable to perform consistently with test-items (on some tasks they could regulate, on other tasks they could not), both groups performed worse on word- or picture-matching trials with weaker semantic associations compared to stronger semantic associations (e.g., ship and van versus ship and yacht), and both groups were equally unable to inhibit irrelevant distractors. The findings propose that general cognitive control and the control processes in the CSC framework (i.e., semantic control) are interrelated, but that a deficit in semantic control results in a greater difficulty with conceptual knowledge. In fact, cognitive control abilities such as inhibition and flexibility are factors that help regulate the retrieval of concepts since inhibition allows inappropriate knowledge to be disregarded or unretrieved, whereas flexibility allows mental reflection about semantic associations (e.g., that a carrot can be related to a basketball somehow), therefore the relationship between semantic control and general cognitive control is not that surprising (Thompson et al., 2018). However, while the processes of semantic cognition outlined in the CSC framework have been well established in patients, research has moved forward to investigate semantic cognition and control processes in neurotypical populations.

Through research with neurotypical adults, the support for two distinct forms of semantic control has emerged (Badre, Poldrack, Paré-Blagoev, Insler & Wagner, 2005; Hoffman, 2018). The first process is called *semantic selection* and occurs when automatic activation of multiple semantic representations creates a competition and requires an individual to select the most relevant representation. For example, the probe ‘carrot’ in a simple semantic word-matching task could initiate parsnip and potato as automatic semantic-matches. If the task then specifically asks you to choose between a parsnip or a potato, a competition between the two automatic alternatives will arise and needs to be resolved in order to complete the task. The second process, called *controlled retrieval*, is contrastively defined as the goal-directed search for relevant associations when automatic activation of semantic knowledge has brought forward irrelevant information. Here, instead of initiating a competition, the two provided alternatives are neither a parsnip nor a potato and a semantic association between the probe and an available alternative needs to be mentally retrieved in order to complete the task since the automatic associations were not an option. Even though controlled retrieval and semantic selection will not form part of the study design in this thesis specifically, the research into the processes have revealed that there is a relationship between semantic control, cognitive control, age and semantic knowledge in neurotypical adults, which are all relevant factors to consider when researching semantic cognition in children.

Hoffman (2018) investigated the relationship between ageing, semantic cognition and cognitive control in neurotypical adults. By adopting a between-group design, Hoffman compared performance differences in semantic knowledge, semantic selection, controlled retrieval and cognitive control in young (18 to 31 years) and old adults (61 to 91 years). To assess semantic knowledge, the participants faced two different types of forced choice word tasks where they had to distinguish a real word from a nonsense word and select synonyms for probe words. The semantic knowledge tasks allowed the researcher to determine participants' vocabulary skill. Each participant also underwent four assessments to test non-semantic cognitive control. Visual attention and task switching was measured with the Trail-Making Task, response inhibition was measured with the Eriksen Flanker Task, cognitive flexibility was measured with the Wisconsin Card Sorting Test (WCST), and mental switching and inhibition was measured with a Verbal Fluency test. To measure controlled retrieval and semantic selection, Hoffman adopted previous forced choice semantic tasks devised by Badre and colleagues (2005). Controlled retrieval was measured through a semantic word-association task where participants were forced to select a word-alternative most related to a word-probe. Here some trials had weakly associated word-targets and other trials had strongly associated word-targets. The different target relationships affected the semantic control requirements and weakly associated word-targets were hypothesised to require more controlled retrieval than strongly associated word-targets due to the lack of an automatic semantic relation for the weakly associated word-target. Semantic selection was measured through a feature word-association task where participants had to pick a word-target that shared a common feature with a word-probe. Here word-targets either had a featural and semantic association to the probe-word (e.g., tar – coal) or only a featural association (e.g., ivy – jade). To induce higher semantic selection demands in the featural-only association trials, one of the distractor-words was a strongly-associated related word (e.g., ivy – league). Three main findings were established. First, older adults were worse at semantic selection but better at controlled retrieval compared to younger adults. Second, older adults had significantly broader semantic libraries compared to younger adults. Third, there was a strong positive correlation between semantic selection performance and non-semantic control from the Wisconsin Card Sort Test for both age groups, but older adults were less accurate on the WCST. The findings suggest that controlled retrieval abilities increase with age whilst semantic selection abilities decrease with age. On top of this, the broader semantic libraries could be a factor in how adults retrieve and select associations. As older adults have a more vocabulary skill this could impede their performance to select the word-alternative with the featural association when a semantic word-distractor is present, but similarly the broader vocabulary knowledge could help retrieve weak associations in the controlled retrieval task. The correlation with the WCST is additionally interesting as it suggests that semantic control has a relationship with non-semantic cognitive flexibility. The worse performance in the WCST for older adults is further supported by the fact that healthy ageing typically results in a natural deterioration in cognitive control (Sanchez-Benavides et al., 2010). However, no other cognitive control tasks correlated with the performance on the semantic tasks, which proposes that semantic control and most cognitive control abilities are separate mechanisms in typical ageing. Overall, the involvement of semantic knowledge, cognitive flexibility and age in Hoffman's study suggests a regression in semantic selection and a progression in controlled retrieval for neurotypical adults. This therefore implies that a progression for semantic selection and a regression for controlled retrieval are likely to take place in development or early adulthood.

So far the research findings and theories about semantic cognition have been based on adults. The past research has identified that semantic representations, control processes and general semantic knowledge all impact semantic cognition (e.g., Jefferies & Lambon Ralph, 2006; Thompson et al., 2018; Hoffman, 2018), yet no attention has been made to identify semantic cognition according to the CSC framework in developing populations. Since children are developing, their conceptual processing and cognitive control abilities may not be fully matured and could consequently lead to presenting similar disadvantages in semantic cognition as neurotypical older adults and patients with brain damage do. However, children may also present new behavioural disadvantages, or advantages, to semantic cognition that have not been exposed yet. Therefore, careful consideration should be taken before applying the adult-theories of semantic cognition to developing populations, and research concerning semantic cognition and children should be encouraged to corroborate or refute the CSC framework in younger populations. Developmental research into semantic cognition can also further identify the maturation of semantic control and semantic representation, and how to facilitate conceptual processing in children by understanding what advantages and disadvantages developing populations have in relation to the processes of semantic cognition. The next section will be presenting research concerning children's concept development, semantic association skills and cognitive control abilities.

1.2 Concept Development, Semantic Associations and Cognitive Control in Children

Research has found that infants are proficient at recognising common objects and are able to deduce semantic associations though preferential looking (Arias-Trejo & Plunkett, 2009; Bergelsen & Swingley, 2012). Whilst concept knowledge is already present in infants, it typically gets richer with age. In fact, going to school has shown to play an important part in the development of concept knowledge because the structure of language used by teachers helps pupils to mentally reflect and infer semantic associations between concepts (Larraín, 2016). Indeed, when teachers introduce new knowledge through instructions and definitions it allows children to add to, and manipulate, their existing knowledge (Larraín & Haye, 2014). For example, if a child learns that vegetables are nutritious, they can allocate that information to carrots specifically because carrots are a type of vegetable, thereby expanding their semantic networks. However, even though education helps to grow children's concept knowledge, the semantic networks of 8- to 12-year-olds are still less connected and smaller compared to 17- to 87-year-olds (Zortea, Menegola, Villavicencio & Fumagalli de Salles, 2014). These studies suggest that while there is room for children's semantic knowledge to progress, children's available semantic knowledge may influence how well they can associate between concepts; if a child knows a concept, they can associate it with other concepts, which is similar to the representational abilities of patients with SD. Interestingly, a recent study managed to draw a parallel between conceptual knowledge in pre-school children and patients diagnosed with SD. By investigating drawings of six animate and six inanimate concepts from three groups of children (47-month-olds, 63-month-olds and 81-month-olds) and three groups of patients with varying in SD severity (most advanced, moderately advanced and least advanced), researchers found that that children's conceptual knowledge progression is similar to the regression of conceptual knowledge in SD (Pozueta et al., 2020). In the study, the researchers identified that while the number of appropriate concept features increased with age for children, the number of features decreased with severity in the patient groups. For example, the concept envelope would be pictured as a rectangle for 47-month-olds, as a rectangle with a flap for 63-month-olds, and as a rectangle with a flap and seal for 81-month-olds. At the same time, the study revealed that the 47-month-olds and the participants in the most advanced SD group were matched on the number of features that they could produce for a given concept, whereas the 81-month-olds produced the most features overall but not significantly more than the moderately advanced and least advanced patient groups. These findings suggest that six-year-old children have more concept knowledge compared to their younger peers, which could be due to them already attending to school. However, the findings also suggest that children younger than seven years still only have the conceptual knowledge comparable to patients with SD, which could imply that young children could have deficits in semantic representations and memory. Yet, while the deficits in SD are due to a loss of knowledge, children are more likely to suffer from a lack of knowledge. In terms of the CSC framework, this could mean that children's hub and spoke connections are not fully matured.

On top of establishing differences in concept knowledge for children younger than seven years, research has investigated semantic network differences within older school-aged children. In a study adopting a semantic fluency task, Sauzéon and colleagues (2004) managed to identify that age positively increased semantic knowledge stores and the availability of semantic associations. In the semantic fluency task, children aged 7- to 16-years-old were required to produce as many words as possible in 60 seconds from one category (supermarket). The findings revealed that older children were able to significantly produce more diversity in terms of the types of products found in a supermarket (e.g., including 'household items') and they significantly produced more words within the different product categories. The findings further confirm that older children have more semantic knowledge and a better grasp on semantic associations compared to their younger peers. Through having these increased abilities, older children may suffer less from a lack of knowledge compared to younger children. However, even though older children may have an advantage due to better concept knowledge, cognitive control also plays a role in the retrieval of semantic representations according to the CSC framework.

Cognitive control is a recognised deficit in typically developing children and adolescents due to an underdeveloped prefrontal cortex (Thompson-Schill, Ramscar & Chrysikou, 2009) and, as children age, cognitive control abilities gradually get better until the prefrontal cortex is fully matured at 24 years of age (Arain et al., 2013). As mentioned previously, cognitive control is typically separated into working memory, inhibition and flexibility (Miyake et al., 2000) and even though studies have not investigated the CSC framework directly in children, researchers have explored how cognitive flexibility and inhibitory control interacts with regulating semantic associations.

In an attempt to understand how children start to self-direct their own actions without parental prompts, Snyder and Munakata (2010) investigated whether switching in a cognitive flexibility task correlated with switching between categories in a semantic fluency task in 5-year-olds. For the study, cognitive flexibility was measured with the Flexible Item Selection Task (FIST) where children were presented with three cartoon pictures that could be matched in two separate pairs based on colour, size or shape. For example, if the stimuli were a small yellow boat, a big yellow boat and a big red boat, the shape boat is irrelevant and to successfully select two pairs of pictures that go together the participants would have to select the two yellow boats (colour) and two big boats (size) in two separate picture pointing movements. The successfulness of the FIST was additionally only measured by the number of correct responses in the second selection after the first selection was successful. The researchers found that switching in the FIST positively correlated with switching during the verbal fluency task, which suggests that cognitive flexibility helps to regulate the retrieval of more diverse responses in children's semantic networks. In terms of the research in adult semantic control, these findings imply that children recruit cognitive flexibility when dealing with semantic knowledge and could therefore also be recruiting semantic control to perform the semantic fluency task since semantic selection correlates with adult switching ability on the WCST. However, the link between cognitive control and semantic associations in children is not exclusive to cognitive flexibility as researchers have additionally found that inhibitory control significantly interacts with how well children can select semantic associations.

Thibaut and colleagues (2011) investigated whether the inhibitory control ability a child has can mediate their performance on a semantic analogy-making task with semantic and featural distractors. For the semantic analogy-making task, four- to five-year old children were presented with a black and white picture pair scenario that they had to reproduce with a picture-probe and a picture-alternative. These picture pair scenarios were two pictures that semantically fit together based on a shared event, for example a bird and a birdcage. The children had four alternatives to choose from to pick the picture-target for a given picture-probe. Two of the alternatives would be unrelated distractors, one alternative would have a semantic or a featural relation to the probe (depending on the trial condition) and the remaining alternative would be the target. In the example where the bird and the birdcage was the picture pair scenario, children had to make a similar paring by selecting a picture of a fish tank for a fish (picture-probe) and exclude pictures of a chair (unrelated distractor), an axe (unrelated distractor) and a body of water with a fisherman (semantic distractor). The semantic relationship of the probes and targets were additionally manipulated by sometimes being weak (e.g., pig and bucket) and sometimes being strong (e.g., tennis ball and racket). As well as completing the semantic analogy-making task, the Day-Night task was used to measure inhibitory control. For the Day-Night task the children were required to consistently say “night” to a picture of a sun and “day” to a picture of a moon. The Day-Night task, although simple, is able to capture children’s ability to inhibit the intuitive response, sun equals day and moon equals night, when a correct response in the game is actually the opposite. The findings showed that children who were poor at inhibiting their responses on the Day-Night task (five or more mistakes) were significantly worse at pairing weakly associated picture pairs compared to children who performed at ceiling on the Day-Night task (three or less mistakes), but no performance differences were found in the strongly associated analogy trials. The results suggests that pre-school children are better at pairing semantically weak pictures together if they have better inhibitory control, but that their inhibitory control maturity mediates their conceptual knowledge to replicate a weak association. This finding with pre-schoolers suggests that children’s ability to infer semantic relationships uses a different type of cognitive control compared to adults who rely on cognitive flexibility. However, the researchers did exclude seven participants who had average inhibitory control abilities due to adopting a between-groups design in the analysis. Therefore, inhibitory control cannot directly be linked to semantic association skills fully because the study provided no explanation for how the average Day-Night participants performed in terms of the weak or strong semantic associations. Moreover, the study additionally found that trials with semantic distractors were more inaccurate compared to trials with featural distractors, and the difference in accuracy remained the same for both age-bands. These additional findings further confirms that semantic distractors place a higher cognitive load compared to featural distractors but age differences in performance are not evident in 4- and 5-year-olds. The added cognitive load for the trials with semantic distractors could be the requirement of semantic control, however this is purely a postulation as semantic control has not been investigated explicitly in children. Nevertheless, the findings from the research investigating the role of cognitive flexibility in semantic fluency (Snyder & Munakata, 2010) and the role of inhibitory control in creating semantic associations (Thibaut, Vezneva, Gérard & Glady, 2011) suggest that children, just like adults, recruit cognitive control abilities to make sense of semantic knowledge.

In the world of developmental research, semantic cognition has not been investigated specifically, but studies do provide findings for certain mechanisms pertinent to semantic cognition. First, research suggests that older children have more semantic knowledge and readily available semantic associations compared to younger children. Second, there is a relationship between cognitive flexibility and semantic knowledge in pre-schoolers. Third, and last, there is a relationship between the ability to infer semantic associations and inhibitory control in pre-schoolers. Due to these three findings, there is a motive to investigate semantic cognition in children with the theoretical knowledge of the CSC framework in mind. The current project will specifically adopt an investigation into the control mechanism of semantic cognition (i.e., semantic control) as a starting point to see if this even is a process in development. Understanding how children fit into the CSC framework and understanding more about the factors affecting conceptual development through semantic cognition allows researchers and educators to construct more specific techniques to facilitate conceptual learning in typical and atypical developing populations.

1.3 The Current Project

The current project aims to investigate whether semantic control can be identified in school-aged children and whether chronological age, vocabulary knowledge or cognitive control can predict semantic control abilities in development. The study is heavily based on transforming Hoffman's project into a child-friendly experiment to understand semantic control in children without specifically separating controlled retrieval and semantic selection from each other. A taxonomic 2 alternative forced choice semantic picture matching task (SPM) was devised to elicit children's semantic control abilities. The SPM required participants to match a picture-probe to the most related picture-alternative. The semantic associations between the picture-probes and picture-alternatives were manipulated in four different trial-types to induce different semantic control requirements. The difficulty of discriminating the picture-target will determine the degree of semantic control that is needed in order to regulate the retrieval of semantic associations.

The SPM was programmed with a mouse-tracking software where geometric mouse-tracking data could be captured for the task. The added geometric mouse-tracking measures allow for further unconscious measures of semantic control to be analysed and these measures have been identified as being sensitive enough to compute the effects of semantic control in adult conceptual processing (Hindy et al., 2009). The use of picture-stimuli was additionally adopted as it cancels out any individual variations in reading skill and it allows for the pictures to control the need to compose mental representations for the concepts (semantic representation).

While semantic control was measured with the SPM, vocabulary knowledge was measured with the third edition of the British Picture Vocabulary Scale (BPVSIII; Dunn & Dunn, 2009). The current and previous editions of the BPVS are standardised receptive vocabulary tests that are reliable in measuring receptive vocabulary in British children. On top of this, there were two measures of non-semantic cognitive control; an inhibitory control measure and a cognitive flexibility measure. These two processes of cognitive control were chosen because past studies with adults and children have found that they relate to how well participants can retrieve semantic associations (Hoffman, 2018; Thibaut et al., 2011; Snyder & Munakata, 2010). Inhibitory control was measured with a Go/No-Go task and cognitive flexibility was measured with a Wisconsin Card Sort Task (WCST). Both the Go/No-Go and WCST have been used successfully with school-aged children before (Cragg & Nation, 2008; Solomon et al., 2009).

Three hypotheses were proposed for the current project based on previous research. The first hypothesis concerns participants' performance on the SPM in terms of the different requirements of semantic control. The within-task control manipulations of the SPM are hypothesised to generate differences in accuracy, response duration and geometric mouse-tracking scores. Poorer performance on all measures is expected in the trial-types that have competing alternatives with very high semantic relations and very low semantic relations to the probe compared to the trial-types where there is a clearer separation between the target and distractor. The first hypothesis is based on the adult and child research that have found varying semantic associations impact performance (Hoffman, 2018; Thibaut et al., 2011).

The second hypothesis concerns chronological age and vocabulary knowledge. Here it is hypothesized that chronological age and vocabulary knowledge will help predict semantic control performance in the SPM. If a child is older or has a better vocabulary, they will be faster at selecting the most semantically related picture-alternative and have more direct geometric mouse-tracking performance compared to children that are younger or have worse vocabularies. The second hypothesis is based on the findings that age and semantic knowledge has affected semantic selection and controlled retrieval ability in adults (Hoffman, 2018).

The third hypothesis concerns cognitive control and anticipates that performance on the Go/No-Go task and WCST will predict performance on the SPM. If a child has better cognitive control abilities, they will be faster at selecting the most semantically related picture-alternative and have more direct geometric mouse-tracking performance compared to children who have worse cognitive control abilities. The third hypothesis is based on the findings that inhibitory control in pre-schoolers affected alternative selection for weak associations (Thibaut et al., 2011), that WCST performance affected semantic selection ability in adults (Hoffman, 2018), and that cognitive flexibility performance in pre-schoolers correlated positively to performance on a semantic verbal fluency task (Snyder & Munakata, 2010).

2. Methods

2.1 Participants

Eighty-three children (43 male; 40 female) were recruited and sampled for the study. An opportunity sampling method was used to recruit participants through two channels; (i) through a pre-existing database of research-active families at the Bristol Cognitive Development Centre (BCDC), University of Bristol, and (ii) through primary schools located in Bristol, London and Surrey, UK. Children who were recruited through the database participated at the BCDC (n = 54). At the BCDC parents were reimbursed £5 for their travel costs and children were given a little toy as a thank-you gift. Children who were recruited through schools participated in their own school (n = 29). School testing took place in a quiet space on the school premises where the research activity could be readily overseen and overheard. School testing did not offer travel reimbursement or a thank-you gift.

All participants were blind to the experiment outcome until the end of the study. A parent or guardian filled out a consent form and a background questionnaire on behalf of every child prior to participation. Participants were excluded if they currently had a significant sensory impairment (namely non-corrected impaired vision or hearing loss), if they were diagnosed with a developmental disorder affecting language and/or cognition, or if they were not aged between 6 to 13 years inclusive. Two children were excluded based on having developmental disorders (Attention Deficit Hyperactive Disorder and Autism Spectrum Disorder) and, even though it was not an original exclusion criterion, a third child was excluded upon reflection based on having significant gaps in education due to epilepsy. The remaining 80 typically developing children (42 male; 38 female) were aged between 6 years; 1 month to 12 years; 0 months ($Mean_{age}$ = 8.81 years, SD_{age} = 1.43 years). Figure 1 shows that the sample has an uneven distribution of age by predominantly consisting of 7-year-olds. Ethnicity was also skewed with 74 participants being white (69 British-White). Of the remaining six participants, two preferred not to disclose ethnicity, one was British-African-Caribbean, one was British-Indian-Pakistani-Bangladeshi, one was British-Latin, and one was British-Mixed.

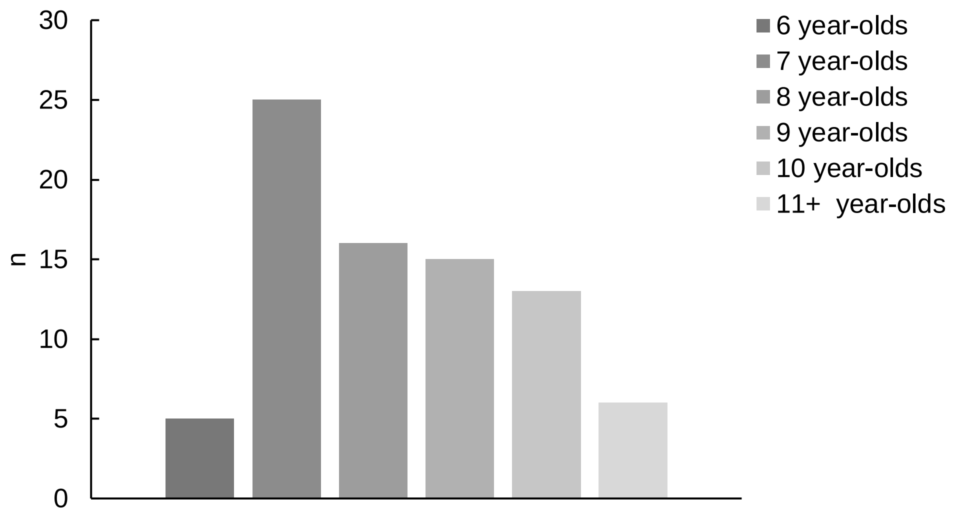

Figure 1. Number of participants per age-band in years.

Table 1 portrays further sample demographics and highlights that the 80 participants were predominantly monolingual English speakers (89%) from educated families (84% of children had at least one parent with a degree-level qualification). Further parental and guardian report suggested that there were several concerns regarding early language production and pronunciation (14%), however, all parental concerns had been resolved before five years of age with or without the help of a Speech and Language Therapist.

Table 1.
Demographics of Study Sample.

Factor	Sample
Highest qualification held in participant household	
n	80
Left senior school around 16yrs (with English, Maths GCS's and extra qualification)	1
Left college (after completing a vocational/technical course e.g., HND)	7
Left college (with A-levels or extra qualifications such as an Access course)	3
Left University (with UG degree)	20
Left University (with PG degree)	47
Other Further Education	2
Participant Language Use	
n	80
English monolingual	71
English bilingual	8
English trilingual	1
Age when participant started speaking	
n	80
4-15 months	59
18-30 months	7
No response/recollection	14
Parental concerns about Early Language Development	
n	80
Not worried	68
Worried	11
No response	1
Speech and Language Therapist Interventions	
n	80
No Interventions due to absence of concern about Early Language Use	67
No Interventions despite concerns about Early Language Use	3
Interventions due to Early Language Concerns	8
Interventions despite no concern about Early Language Use	1
No response	1

Note.

2.2 Apparatus and Materials

Five of the seven tasks in the study were administered electronically. These tasks were the semantic picture matching task (SPM), the Go/No-Go task, the Wisconsin Card Sort Task (WCST) and two separate Mouse clicking games. The electronic equipment used to administer the five tasks were a 19” Iiyama ProLite B1906S screen, a USB Dell 11D3V hand-held computer mouse, a QWERTY keyboard (Kensington ValuKeyboard 1500109), and a Dell computer or laptop operating Windows 7. The remaining two tasks, the British Picture Vocabulary Scale Edition III (BPVSIII) and the Bishop’s Reaching Task, were manually administered.

2.2.1 Semantic Picture Matching Task

A novel computerised taxonomic semantic picture matching task (SPM) was created to measure semantic control. The SPM was programmed in MouseTracker (Freeman & Ambady, 2010) in order to collect geometric cursor movement measures. Each trial in the SPM consisted of three different pictures; the picture-probe was displayed centrally at the bottom of the screen and the two picture match alternatives were located in the top-right and top-left hand corners surrounded by a black border. The three pictures in a trial were all shown at the same time once the trial had begun. Only one response could be recorded in every trial by moving the computer mouse and clicking on one of the two picture match alternatives. Cursor speed was set to five (maximum speed is 20) to allow for better handling of the cursor.

2.2.1.1 Stimuli development. 150 pictures were used to make the trials for the SPM and six steps were taken to determine the picture stimuli. First, test-based age-of-acquisition data from Brysbaert and Biemiller (2016) were used to generate a list of 30 probe words with the criteria of being a noun acquired before the age of 6 years. Three probe words were included despite not meeting the age criteria (i.e., Rucksack, Trainers, and Budgie). The three exceptions were allowed because the rated equivalent words in American English (i.e., Knapsack, Sneakers, and Parakeet) met the early acquisition criteria.

The second step was to generate four possible word-alternatives for every chosen word-probe based on their semantic similarity by taxonomic relations. The word-alternatives were organised in a hierarchical manner; there was a *very proximal* alternative, a *proximal* alternative, a *distal* alternative, and an *unrelated* alternative for each word-probe. Every word-alternative was generated by researching multiple datasets from past studies of semantic relatedness (e.g., De Deyne, Navarro, Perfors, Brysbaert & Storms, 2018; Landrigan & Mirman, 2016; Maki, McKinley & Thompson, 2004).

As the third step, Latent Semantic Analysis (LSA) ratings were recorded for the pairing of each word-alternative in relation to its respective word-probe to ensure the four pairings differed in semantic relatedness (very proximal, proximal, distal, and unrelated). LSA ratings were computed using the pairwise comparison function from an LSA Database Website and provide a mathematical value to how related words are (Laham & Steinhart, 1998). Each probe-alternative word pair was submitted into the pairwise comparison function that searched through a topic space called *General Reading up to First Year College* where the maximum number of factors was 300. Since LSA values were established for every probe to alternative word pair, a one-way Analysis of Variance (ANOVA) was performed to investigate whether the collected LSA values significantly differed by the alternative type (i.e., very proximal, proximal, distal, and unrelated). The ANOVA uncovered significant LSA differences between the alternative types (F (3, 115) = 24.26, $p < .001$) and a Tukey post hoc test revealed that all alternative types had significant LSA differences to each other ($p \leq .009$), except for the proximal and distal LSA values (p =.825).

It is important to note that the LSA ratings are based on how often words appear in text together, which means that the semantic similarity ratings obtained through the LSA database can be more thematically led than taxonomically led (Mirman, Landrigan & Britt, 2017). Therefore, as a fourth step, a pilot study was performed to obtain adult taxonomic ratings of the probe to alternative word pairs. Thirty undergraduate and postgraduate students ($Mean_{age}$ = 21.07, SD_{age} = 4.54) rated 120-word pairs with an 11-point Likert scale. This pilot study was programmed with Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2019). The participants were instructed to rate the word pairs based on how related they were, taking into account the category the words belonged to and the standard functions that the nouns had (e.g., a hairbrush is used for brushing hair, a paint brush is used for painting a canvas). A response of 0 represented 'Not Related' and a response of 10 represented 'Highly Related'. To understand whether the adult taxonomic ratings differed by alternative type, a one-way ANOVA was performed on the average adult taxonomic rating for every probe to alternative word pair. The ANOVA uncovered significant adult rating differences between the alternative types (F (3, 116) = 202.1, p < .0001) and a Tukey post hoc test revealed that all alternative types had significant differences in adult ratings in relation to each other ($p \leq .048$). The pilot confirmed that all the alternative types varied by taxonomic semantic similarity and strengthened the use of the stimuli.

After finalizing the word list with adults, the fifth step was to assess its suitability for children using exposure ratings to ensure that children would have encountered the words and likely 'know' them. Developmental written and televised exposure ratings were obtained for every word. The written exposure ratings originated from the Children's Printed Word Database (Lovejoy, 2003), where the number of times a word has been printed in a book directed to 5- and 9-year-olds is pooled. A one-way ANOVA revealed no significant differences in written exposure between the probe, very proximal, proximal, distal and unrelated groupings (F (4, 135) = .195, p = .94). The televised exposure data was obtained from SUBTLEX UK (Van Heuven, Mandera, Keuleers & Brysbaert, 2014), where the number of times a word is used in CBBC and CBeebies subtitles is calculated. A one-way ANOVA revealed no significant differences in televised exposure between the words in the probe, very proximal, proximal, distal and unrelated groupings for both CBBC and CBeebies subtitles (F (4, 143) = .801, p = .526, F (4, 143) = .782, p = .539 respectively). The collected written and televised exposure data further suggested that the stimuli was fit for children by showing that children aged six years and above were likely to have been exposed to the words in written and spoken forms. The full list of word-stimuli for the SPM can be found in Appendix A.

The final step was to convert all the words into pictures. The pictures for the SPM were obtained from Google images by searching for the words in the Google search engine. All pictures had a white background and the subject (i.e., the noun) was pictured alone. The pictures were then resized into images that measured 500x300 pixels through a picture resizing website (Online Image Resizer, 2015).

2.2.1.2 Design. Four trial-types were generated from the picture stimuli to measure semantic control in the present study. The four trial-types created two contrasts. The first contrast (referred to as the semantic knowledge contrast) compared two trial-types where the picture-target was either proximally related (SK-proximal) or distally related (SK-distal) to the picture-probe. The picture-distractors in the SK-proximal and SK-distal trials were always the unrelated picture-alternative to the picture-probe. Critically, the semantic relation between the picture-probe and the picture-target determined the semantic control requirements in the semantic knowledge contrast (see Figure 2a). The second contrast (referred to as the semantic control contrast) compared two trial-types where the picture-target always had a very proximal semantic relation to the picture-probe. The picture-distractors in the semantic control contrast either had a proximal semantic relation to the picture probe (SC-proximal) or a distal semantic relation to the picture-probe (SC-distal). Critically, the semantic relation between the picture-probe and the picture-distractor determined the semantic control requirements in the semantic control contrast (see Figure 2b).

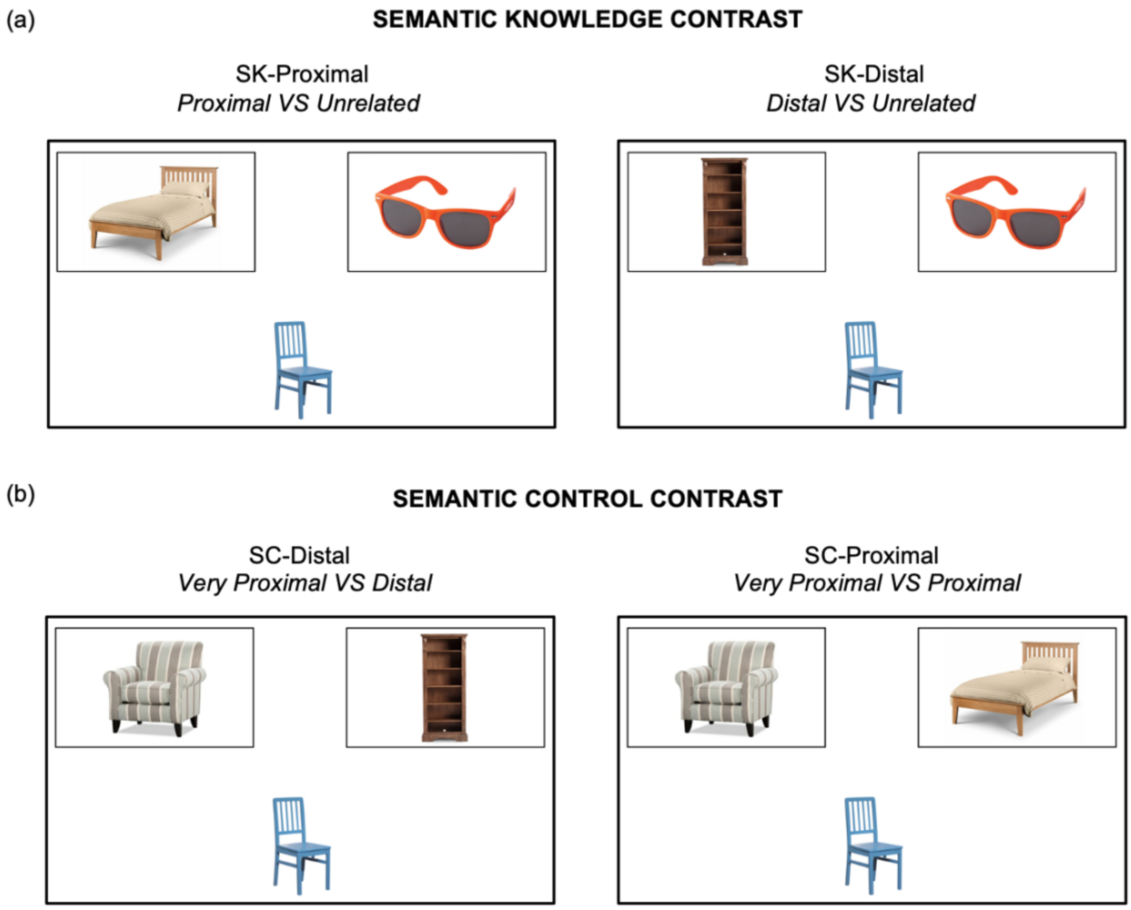

Figure 2. Sample of the four trial-types used in the SPM with one set of pictures.

By adopting the semantic knowledge and semantic control contrasts, each picture-probe was used twice to create 60 trials in the SPM. To avoid excess repetition due to recycling the picture-probes, two sets of 30 trials were compiled (set A and B). Each set contained identical picture-probes, but the trial-type for each picture-probe differed between the sets. For example, if a trial with a blue chair picture-probe was a SK-Proximal trial in set A, then the same probe (i.e., blue chair) would be used as a SC-Distal trial in set B. However, if the trial was SK-Distal in set A, the trial would be SC-Proximal in set B.

There were 15 trials of each trial-type across the 60 trials. Each picture-probe was quasi-randomly assigned two trial-types and each trial-type of a picture-probe was quasi-randomly assigned to set A or B. Care was taken to ensure that each set contained an even distribution of trial-types (7 or 8) and that the order of trials in each set were randomised in a fixed order.

To avoid additional bias in the mouse-tracking measures arising from the geographical location of the picture-target, two further constraints were introduced to vary the position of the picture-target across trials. Firstly, the position of the picture-target was quasi-randomly assigned to the right-hand or left-hand corner of the screen. Secondly, two *mirrored* versions of the sets (set A mirrored and B mirrored) were created that exclusively flipped the geographical position of the picture-target horizontally to the opposing side of the screen compared to the non-mirrored equivalents. Figure 3 demonstrates how a SK-proximal trial (Figure 3a and 3c) and a SC-distal trial (Figure 3b and 3d) changes depending on what set it is in.

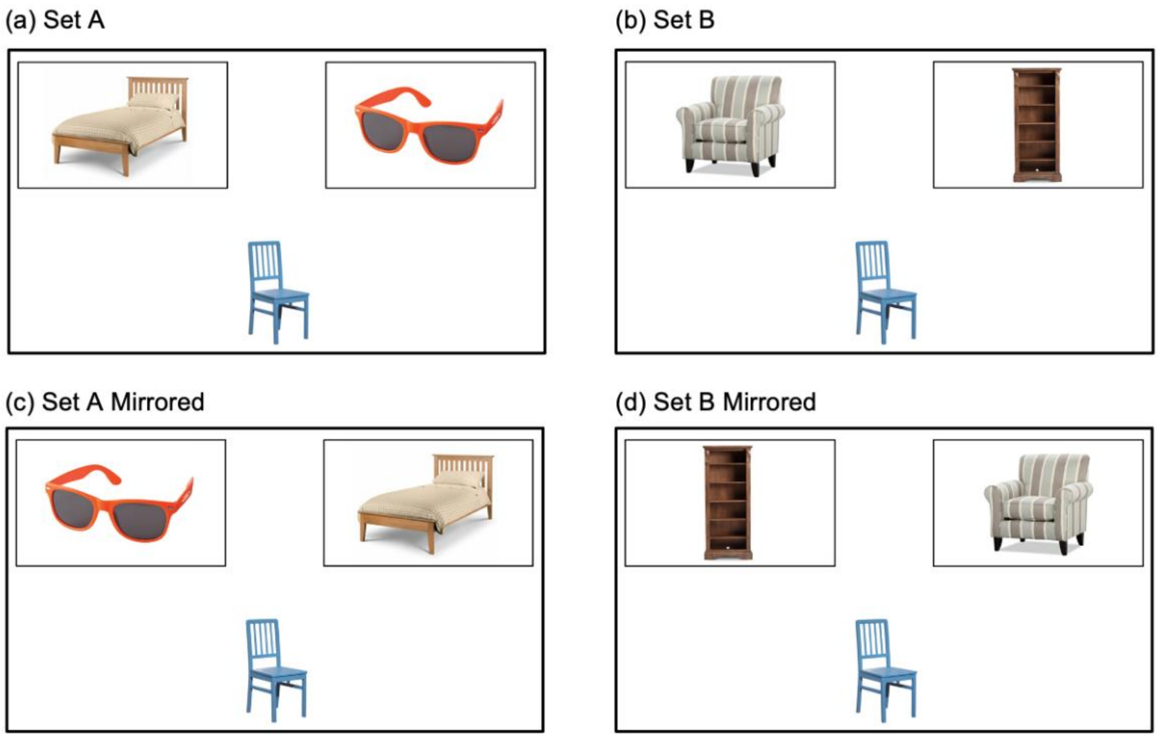

Figure 3. Sample of the picture-alternative location changes in unmirrored sets (a and b) and mirrored sets (c and d) in the SPM.

The two extra mirrored sets meant that participants either completed set A and set B or completed set A mirrored and set B mirrored. To eliminate any performance bias of having a particular set first, four presentation orders were implemented. Participants either received set A then set B, set B then set A, set A mirrored then set B mirrored, or set B mirrored then set A mirrored.

On top of the 30 trials in each set, there were eight instruction screens and three practice trials before the test trials. Any pictures used in the practice trials and instructions were found separately on Google Images beyond the 150 pictures used in the test trials.

2.2.2 Bishop's Reaching Task

Bishop's reaching task was used to determine participants' hand preference (Bishop, Ross, Daniels, & Bright, 1996). The task requires participants to pick up cards that are laid out on a flat surface and the experimenter notes what hand (left or right) was used for the card pick-up. The materials consisted of 21 laminated A7 cards that had different coloured shapes on them. The shape was either a heart, circle or triangle, and the colour variations were black, blue, green, pink, red, white and yellow.

2.2.3 BPVSIII

The third edition of the British Picture Vocabulary Scale (BPVSIII) was administered to capture receptive vocabulary knowledge. The BPVSIII manual, BPVSIII testing book and BPVSIII record forms were used to administer the task. The task requires children to match a word, spoken by the experimenter, to one of four coloured pictures that are all displayed on one page in the BPVSIII testing book. Every 12 pages of the BPVSIII testing book are referred to as a set and each set is directed to a specific age or age-range.

2.2.4 Go/No-Go

A Go/No-Go task was used to test inhibitory control. The task was set up as a fishing game inspired by previous Go/No-Go tasks used with pre-school and school aged children

(Cragg & Nation, 2008; Wiebe, Sheffield & Epsy, 2012), and was programmed with Gorilla Experiment Builder (Anwyl-Irvine et al., 2019). To measure inhibitory control, participants were required to respond to Go trials (indicated by a fish) and withhold a response to the No-Go trials (indicated by a shark). Five different cartoon pictures of fish and two different cartoon pictures of sharks were found through Google images to be used as the stimuli. In total there were 58 trials: 10 practice trails and 48 test trials. The practice trials were divided into two separate blocks of 5 trials. The first practice block contained exclusively Go trials and the second practice block contained exclusively No-Go trials. The 48 test trials were programmed as one block and randomized into a fixed presentation order where 75% were Go trials and 25% were No-Go trials.

2.2.5 Mouse Clicking Practice Games

Two different mouse clicking practice games were used to give children some practice on using a hand-held computer mouse before they started the experiment tasks and were administered together as one game.

2.2.5.1 Dot clicking game. A computerised dot clicking game was created using Gorilla Experiment Builder (Anwyl-Irvine et al., 2019). The game consisted of clicking on a series of coloured dots, one at a time, that appeared on-screen. Once a dot was clicked, the next appeared without delay. The dots were one of eight different colours (black, green, purple, blue, orange, pink, yellow, red), one of 12 different sizes (ranging from occupying 9% of the screen to .25% of the screen) and in one of 12 different screen locations. In total there were three demonstration trials, three practice trials and 24 test trials.

2.2.5.2 Bubble pop game. The second mouse clicking practice game was hosted on the Let's Go Learn website (Let's Go Learn, 2020). The game was called the '1 minute Mouse Practice Activity' and participants were required to click on as many bubbles as they could in 60 seconds. The bubbles were transparent and appeared one at a time on a blue background (mimicking a sky). In October 2019, Let's Go Learn updated the game and made minor alterations. The update meant that after 10 successful popped bubbles, a cheering crowd appeared and delayed the appearance of the subsequent bubble. No data were collated from the task due to the update.

2.2.6 Wisconsin Card Sort Task

A Wisconsin Card Sort Task (WCST) was used to test cognitive flexibility. The WCST was programmed with Gorilla Experiment Builder (Anwyl-Irvine et al., 2019) and was based on a demo made by PsychToolKit (Berg, 1948; PsychToolKit, 2021). To measure cognitive flexibility, participants were required to match a probe card to a response card based on the type of shape on the cards, the number of shapes on the cards or the colour of the shape(s) on the cards with the help of feedback. Due to the possibility of multiple response options, the participant has to flexibly change matching method if feedback indicates an incorrect match (not choosing the correct response card). The card stimuli were created by colouring four different shapes (i.e., circle, triangle, addition and star shapes) in four different colours (i.e., green, red, yellow and blue) using Microsoft PowerPoint. The feedback stimuli were a cartoon-picture of a green tick and a red cross, which were retrieved from Google images. The WCST consisted of 48 test trials that were presented as six consecutive blocks of eight trials. For blocks one and four a correct match was by the type of shape on the cards. For blocks two and five a correct match was by the type of colour of the shapes on the cards. For blocks three and six a correct match was by the number of shapes on the cards.

2.3 Design

The study adopted a within-subjects design to explore performance on the SPM. Due to the four possible presentation orders, participants were stratified into presentation orders based on their age and sex to balance the presentation orders across participants.

The first hypothesis investigated the performance differences in accuracy, response duration and geometric mouse-tracking measures between the trial-types that vary the picture-target semantic relation to the picture-probe (SK-proximal versus SK-distal) and the trial-types that vary the picture-distractor semantic relation to the picture-probe (SC-distal versus SC-proximal). The independent variable was trial-type and the dependent variables were accuracy (proportion of correct responses as a percentage), response duration (reaction time in milliseconds from the first movement of the computer mouse to picture alternative selection), Area Under the Curve (AUC; the size of the curve) and Maximum Deviation (MD; the largest perpendicular difference between the curve and the most ideal cursor trajectory). Please refer to Appendix B for visuals on how AUC and MD are calculated.

The second hypothesis investigated whether chronological age and vocabulary knowledge could predict response duration and geometric mouse-tracking performance in the SPM. The independent variable here was trial-type and the dependent variables were response duration, AUC and MD. The predictor variables were chronological age, divided into two groups (6- to 7-year-olds and 8- to 12-year-olds) and BPVSIII raw score.

The third hypothesis investigated whether cognitive control ability could predict response duration and geometric mouse-tracking performance in the SPM. The independent variable here was trial-type and the dependent variables were response duration, AUC and MD. The predictor variables were Go/No-Go sensitivity and the proportion of perseveration errors in the WCST post switch trials.

The Bishop's Reaching Task and the Mouse Clicking Practice Games were purely used to generate controls on performance. The data from these tasks did not form part of any hypotheses, but are reported to show the dot clicking speed in milliseconds by chronological age of the sample and the hand preference (right or left) of the sample.

2.4 Procedure

Due to the two recruitment channels, children participated in the BCDC or at their own primary school. The children who participated in the BCDC completed all the tasks in a 40-minute session with one mandatory 5-minute break halfway through the experiment. Parents were invited to accompany their child in the testing room if they wished to do so. The children who participated in their own school were visited twice by an experimenter for two separate 20-minute sessions. The breaks between the first and second visit at schools were between 1 day and 14 days due to school scheduling. The data was collected from June 2019 to March 2020.

In the test sessions, all children participated individually with an experimenter whilst sitting at a table with the computer screen, keyboard and computer mouse in front of them. The order of tasks was kept as constant as possible despite the two different testing places. The pre-break tasks were the Mouse Clicking Practice Games, the SPM set 1, the Go/No-Go and the BPVSIII, and the post-break tasks were the WCST, the Bishop's Reaching Task and the SPM set 2. All children collected a sticker for every task they completed during testing and the tasks were introduced as fun games for the children. All the electronically administered tasks were portrayed in full screen mode.

2.4.1 Semantic Picture Matching Task

The SPM was presented to children as a picture matching game in two separate sets of 30 trials. At the start of each set, the experimenter explained the game alongside screens with written instructions or helpful visuals. The instructions started with a general introduction to the task: "Now we are going to play the picture matching game. For the picture matching game, you are going to see a grey start button at the bottom of the screen. Once you click it, three pictures will appear, two at the top and one at the bottom.". Then the experimenter continued by explaining how to match the pictures to each other: "What I want you to do is click on the top picture that you want to choose as a match for the bottom picture. And I want you to match the pictures based on how related they are.”. The instructions were accompanied by exemplars: “So let’s see some examples to understand what I mean by related. Peppa Pig is related to Daddy Pig because they are from the same family. Daddy Pig is related to Peppa Pig. This is the best match. Peppa Pig is not related to Mickey Mouse because they are from different families. Mickey Mouse is not related to Peppa Pig. This is a bad match.”. In the next screen, another example was provided with verbal instructions and explanation of the semantic match with stimuli closer to the test trials: “This is a picture of an Ice Lolly, the Ice Lolly is related to the Ice Cream because they are both cold foods from the freezer. The Ice Cream is related to the Ice Lolly. This is the best match. The Ice Lolly is also related to the grapes because they are both foods. But this is just an ok match, the best match to the Ice Lolly is the Ice Cream.”. After the examples, the participants were instructed that they would practice with three trials. The practice trials were identical for both sets and were purely completed to acclimatise the children to the task layout. After the practice trials, the experimenter reiterated the instruction that the children had to click on the most related top picture to the bottom picture by picking the best match. The experimenter added that the children should be as quick as they can. Then, the 30 test trials were presented. Each set took 5 minutes to complete.

MouseTracker recorded 12 variables for every trial, however only five measures were used for analysis. Raw AUC (measured in squared units) and raw MD (measured in constant units) were used directly for each trial from MouseTracker. Response duration was calculated by subtracting the initiation time (time to start moving the cursor) from the overall time-length of each trial. Accuracy was calculated by inverting the error score for each trial and calculating the proportion of correct trials per trial-type.

2.4.2 Bishop's Reaching Task

The Bishop's reaching task was administered manually and was introduced as a card game to the participants. At the start of the task, the experimenter divided the 21 laminated cards into seven stacks of three cards and laid the stacks out in a half circle formation in front of the participant, making sure they could reach each stack with ease. The children were subsequently told that in order to complete the card game, they had to find the card the experimenter would utter and place it in a new stack to collect all the cards again. The cards were recalled in a fixed order, but the seven stacks were different for each participant. The number of left-, right-, left-to-right-, right-to-left- and two-handed card pick-ups were recorded on a score sheet. The task took approximately three minutes to complete.

The proportion of pick-ups using the right hand first or exclusively was computed and 0.5 was subtracted from the proportion as per recommendations from Calvert and Bishop (1998). The 0.5 reduction created an index where any participant with a value above zero were classed as having a right-hand preference and any participants with a value below zero were classed as having a left-hand preference. Any participant with a value of zero would be classed as having no preference (although due to the uneven trial-number this was improbable).

2.4.3 BPVSIII

The BPVSIII was administered manually as per the instruction manual. To explain the task to the children, the experimenter told the participants that they were going to play a picture pointing game. The experimenter continued to say "For the picture pointing game I am going to show you some pictures from my big booklet of pictures and say a word. What I want you to do is point to, or say the number of, the picture the word is describing. If you are not sure what picture the word belongs to, then I want you to give me your best guess. If you are not comfortable guessing, then let me know but it is always good to guess". If the participant said they did not know the word or were unsure what picture to pick, they were further prompted to guess on the spot. The children started at a set that was age appropriate in order to establish a basal set (12 correct word to picture matches) and kept doing the task until a ceiling set was reached (eight or more wrong word to picture matches). By matching a word to a picture correctly, the BPVSIII assumes that the participant can verbally and visually recognize the identity of a concept. Participants were given praise and comments with encouragement in line with BPVSIII manual instructions to keep the momentum of the task going. The task was completed once the basal and ceiling sets were established. The task took on average 10 minutes to complete.

2.4.4 Go/No-Go

The Go/No-Go was administered electronically via the preview task option in the experiment builder function on the Gorilla website (Anwyl-Irvine et al., 2019). The task was introduced as a fishing game and the keyboard was brought closer to the participant. The experimenter instructed the participant on how to play the game with the help of visual aids presented across eight instruction screens. The instructions specifically stated that the spacebar on the keyboard should be pressed when a fish appeared onscreen (Go trial) but should not be pressed when a shark is present onscreen (No-Go trial). Then, the participants practised the game with the practice blocks before the presentation of 48 test trials.

Each trial consisted of a 500ms central fixation point screen, followed by a 1500ms central stimulus screen, and then a 1000ms blank screen. The blank screen appeared as soon as a response was given, or as soon as the stimulus screen time limit had passed. Participants could only record a response during the stimulus screen by pressing the spacebar on the computer keyboard. Participants did not receive feedback on their performance. A correct response was recorded when the spacebar was pressed for a Go trial and when the spacebar was not pressed for a No-Go trial. Reaction time in milliseconds, Go accuracy and No-Go accuracy were automatically recorded for every trial. Once the test trials were completed, a data file was downloaded. The Go/No-Go task took on average four minutes to complete.

For the data analysis, the proportion of correct Go trials and the proportion of incorrect No-Go trials were calculated. Sensitivity scores were calculated to estimate the relative accuracy of supressing a response to the No-Go trials. The excel formula “NORMSINV” was used to calculate sensitivity scores by subtracting the z-score values of the Go trial accuracy right-tail p-values from the z-score values of the No-Go trial inaccuracy right-tail p-values (Wiebe, Sheffield & Espy, 2013, p. 1252).

2.4.5 Mouse Clicking Practice Games

The two different mouse clicking practice games were the first tasks that the participants completed and the children that were tested in their own schools did both mouse clicking practice games twice; once at the beginning of visit one and a second time at the beginning of visit two.

The first mouse clicking practice game was the dot clicking game. The dot clicking game was administered electronically via the preview task option in the experiment builder function on the Gorilla website (Anwyl-Irvine et al., 2019). The experimenter first explained how to play the game by using three instruction screens and three demonstration trials. After the three practice trials, the participants were told that they now had 24 dots to click and that when they had finished another mouse clicking game was going to be played. There was no time limit to complete the game and reaction time in milliseconds per test trials was recorded.

The second mouse clicking practice game was the bubble pop game. The bubble pop game was administered electronically on the Let’s Go Learn website (Let’s Go Learn, 2020). The experimenter explained the game verbally before the participant started playing. The participants completed the bubble pop task twice to allow them a chance to beat their first score.

The number of bubbles popped in 60 seconds for each play through was recorded. Both mouse clicking practice games combined took five minutes to complete.

2.4.6 Wisconsin Card Sort Task

The WCST was administered electronically via the preview task option in the experiment builder function on the Gorilla website (Anwyl-Irvine et al., 2019). The task was introduced as a card matching game. The experimenter instructed the participant on how to play the game with the help of visual aids presented across seven instruction screens. First, the layout of the cards on the screen for every trial was explained. Each trial consisted of five cards; the probe card was located centrally at the bottom of the screen and the response cards were located evenly spaced at the top of the screen. The response cards remained the same throughout the game (one red circle, two blue triangles, three yellow additions and four green stars) and the probe card changed for every trial. Next, the different matching methods was demonstrated with a card picturing a blue star. The experimenter showed that the blue star could be matched to the red circle (number), to the blue triangles (colour) or to the green stars (shape). Then, the experimenter explained that a response was recorded by clicking on the desired response card. Lastly, the feedback system was explained. If the match was correct, a green tick appeared in the left-hand bottom corner and if the match was incorrect, a red cross appeared. Feedback would appear immediately after every response and remain on the screen for 700ms. Once the instructions were completed, the 48 test trials begun.

The participants had no time limit to complete the game, but they were encouraged to select the correct response card as quickly as they could. To encourage the participants further, the experimenter would reiterate the feedback to the participants at the start of the next trial (such as saying, “no, we are not matching by shape what can it be?” or “yes, we are matching by shape”) to keep the momentum of the game.

Once the task was completed, the experimenter downloaded a data output file from Gorilla with reaction time in milliseconds and response card selection for every trial. The accuracy and inaccuracy of responses were coded for each trial according to the matching rule of the block. The responses were coded as (i) correct, (ii) an absolute error, or (iii) a perseveration error. A correct response was recorded when the response of a trial corresponded to the matching rule (e.g., when a match by colour was selected in block two). An absolute error was recorded when there was no way the response could have been matched to the probe according to its shape, colour or number of shapes (e.g., three blue stars cannot logically be matched to the single red circle response card). A perseveration error was recorded when a participant continued to match by using the wrong matching method despite receiving feedback that it is the incorrect matching method (e.g., receiving a red cross when matching by colour in block three but trying to match by colour again). All responses from the first trial of every block were excluded from the coding due to the first trial being a *switch trial*. On the switch trials participants were unable to anticipate a correct response to the matching rule of the new block without any feedback.

For the purpose of data analysis, cognitive flexibility was calculated as the proportion of perseveration errors generated from the first four trials after the switch trial (trials two to five) in blocks two to six. Maximum perseveration error was capped to 20 and block one was excluded upon reflection due to acting as a practice block rather than testing block. Only post switch trials were used due to past research suggesting that flexibility is best captured by the performance in the first few trials after a switch trial in switching tasks (Doebel & Zelazo, 2015).

2.5 Data Analysis

To test the hypotheses of this study, all data analyses were conducted using R (R Core Team, 2020). A specific R package, called lm4 (Bates, Maechler, Bolker & Walker, 2015), was used to generate a series of generalised linear mixed-effect models (GLMMs). GLMMs allow for an estimation of random effects to be made, which is ignored when using a repeated measures ANOVA (Wainwright, Leatherdale & Dubin, 2007). Random effects, such as individual variation, is common in developmental samples and are important to incorporate in analyses in order to eliminate potential interference. In addition, GLMMs allow for additive effects from different distributions to be fitted (i.e., binary distributions), therefore data such as age-group can be incorporated with ease. Likelihood Ratio Tests were used to obtain p-values of the models with and without predictors. Likelihood Ratio Tests estimate the probability of the data in the model to find the optimal fit of the data. Assumption testing for the GLMMs were performed and no violations were flagged.

One seven-year-old male was excluded from all analyses due to poor behaviour during testing (deliberately not following instructions). Of the remaining 79, one participant had only completed one SPM set and not completed the BPVSIII due to school illness, another participant did not complete the WCST due to experimenter error, and two participants did not complete the Bishop's Reaching Task due to experimenter error. Despite having partial or uncompleted data, these four participants were included in all the analyses where data had been provided.

The dot clicking task was used as a guide to understand how well children could control mouse movements. Visual data from the task showed that task duration related quite highly with chronological age and that there were several outliers amongst younger children (see Figure 4). No action was taken to exclude the participants that were outliers in the dot clicking game from the SPM analyses as the task was used purely as a practice with the bubble pop game. However, action was taken to exclude trials based on response duration in the SPM.

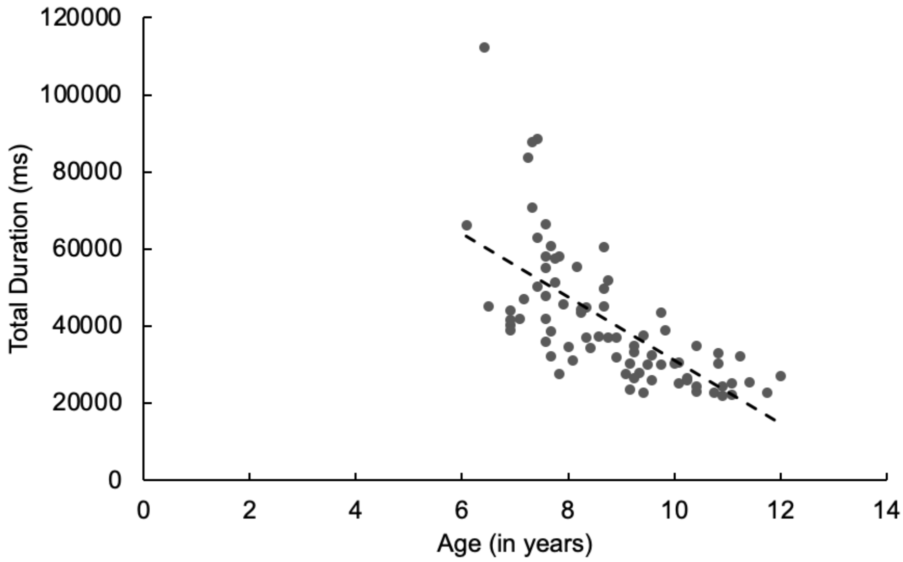

Figure 4. Total duration during the 24 test-trials of the Dot clicking task by age in years (n = 79).

Trials in the SPM that had a response duration over three standard deviations from the mean were removed from the GLMM analyses due to the concern that a very slow response time meant abnormal cursor trajectories. Thirty-four trials were removed from the semantic knowledge contrasts (16 were SK-distal trials) and 36 trials were removed from the semantic control contrast (18 were SC-proximal trials).

3. Results

The results will be presented in four sections. The first section will explore the effectiveness of using MouseTracker with children in regard to their hand preference and the geographical location of the picture-target in the SPM. The second section will explore how accuracy, response duration and geometric mouse tracking performance during the SPM varies by trial-type within the semantic knowledge and semantic control contrasts. The third section will explore how age and vocabulary knowledge predicts SPM trial-type response duration and geometric mouse tracking performance within the two different contrasts. Finally, the fourth section will explore how cognitive control performance predicts SPM trial-type response duration and geometric mouse tracking performance within the two different contrasts.

3.1 Hand Preference and Geographical Location of Picture-Target

In an attempt to understand how hand preference and the geographical location of the target-picture influences children's geometric mouse-tracking scores in the SPM, two separate independent samples t-tests were performed to quantify whether children who completed mirrored sets had significantly different AUC and MD scores compared to children who completed unmirrored sets. Pooled AUC and MD scores from all trial-types was used, and a parametric approach was adopted due to the data passing all appropriate assumptions.

In the overall sample, 58 children were classed as having a right-hand preference (75%) and 19 children were classed as having a left-handed preference. The hand preference in the overall sample reflected a unequal split in the two presentation sets. Children who completed unmirrored sets had a 70/30 percent split of right-and left-handers (M_{AUC} = .48, SD_{AUC} = .28; M_{MD} = .26, SD_{MD} = .12). Children who completed mirrored sets had an 81/19 percent split of right- and left-handers (M_{AUC} = .52, SD_{AUC} = .41; M_{MD} = .26, SD_{MD} = .15). In addition, two participants used their left hand on the computer mouse, one of them were assigned to the mirrored presentation sets and the other to the unmirrored presentation sets.

The independent samples t-test concerning AUC was non-significant with a small effect size ($t(77) = -.55$, $p = .587$, $d = .12$) and the independent samples t-test concerning MD was also non-significant with a small effect size ($t(77) = -.06$, $p = .953$, $d = .01$). The findings suggest that target-picture location does not encourage differences in mouse-trajectories.

3.2 Exploring Semantic Control in the SPM Without Predictors

The first hypothesis seeks to test whether trial-types with competing picture-alternatives that have very low semantic relations (SK-distal) and very high semantic relations (SC-proximal) are less accurate, take longer, and create bigger AUC and MD scores compared to their matched trial-type equivalents that have a clearer separation between the target and distractor (SK-proximal and SC-distal). The first hypothesis was examined using accuracy, response duration and geometric mouse curvature scores from the novel computerised taxonomic SPM task. The analyses that include the response duration and geometric scores discards the inaccurate trials from analysis because the target alternative was not chosen and the geometric scores were not recorded during incorrect trials. Figure 5 shows the mean scores on the SPM for all trial-types.

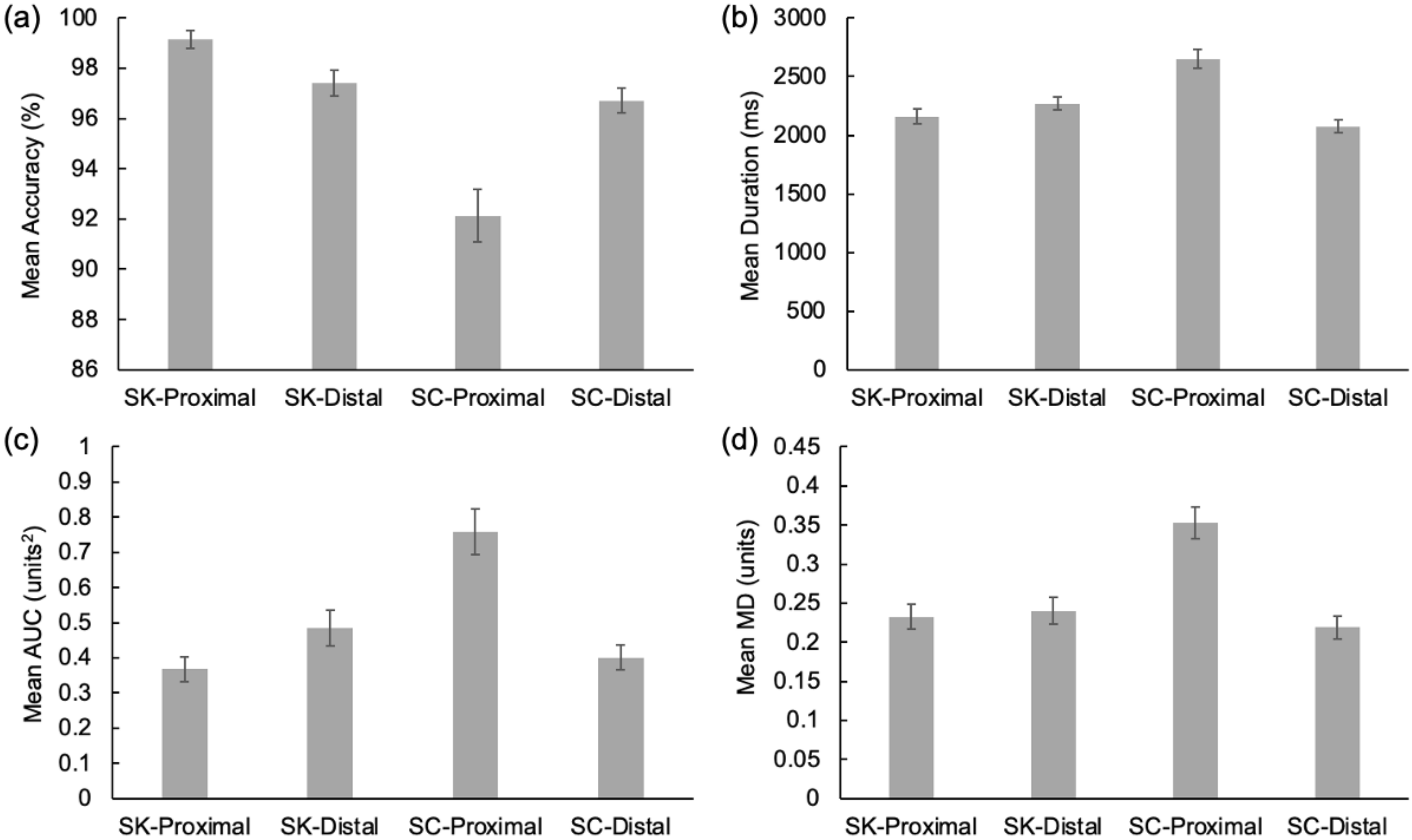

Figure 5. Mean accuracy, response duration, AUC scores and MD scores for the SPM trial-types with standard error as error bars.

Figure 5a shows that the participants were on average the most accurate during SK-proximal trials and the least accurate during SC-proximal trials. Despite the differences between trial-types, children generally performed well on the SPM and all accuracy scores were close to ceiling.

Figure 5b shows that average response duration remained even across all trial-types and children took on average less than three seconds to make their alternative choice. SC-proximal trials took on average the longest time to complete and the SC-distal trials took the shortest time to complete.

Figure 5c shows that SC-distal and SK-proximal trials yielded the smallest average curve sizes (AUC), whilst the average AUC during SC-proximal trials was twice the size. SK-distal trials yielded the second biggest AUC on average.

Figure 5d shows that the average MD for SK-proximal, SK-distal and SC-distal trials were very similar, whereas the average MD for SC-proximal trials was approximately a third bigger compared to the other three trial-types.

3.2.1 Semantic Knowledge Contrast

3.2.1.1 Accuracy. Figure 5a shows that SK-proximal trials were marginally more accurate compared to SK-distal trials. The mean difference was 1.6%. Due to the accuracy scores being non-normal, a Wilcoxon Signed Rank Test was used to test for a statistical difference between SK-proximal and SK-distal trials. A significant difference was found for accuracy scores between SK-proximal and SK-distal ($z = -2.99$, $p < .002$, $r = -.33$). The analysis indicated that SK-proximal trials were on average more accurate than SK-distal trials and that the size of the effect was medium.

3.2.1.2 Response Duration and Geometric Mouse-tracking Scores. Three GLMMs were used to assess the effect of trial-type for the semantic knowledge contrast in terms of response duration and for geometric measures of mouse curvature. Trial-type was entered as a fixed factor and participant was entered as a random factor.

Three Likelihood Ratio Tests were computed to compare the GLMMs to random-effects only models (GLMMs with no fixed factor). The model comparisons concerning response duration and AUC were significant (Duration Δ in $AIC = 7$, $X^2(1) = 9.36$, $p = .002$; AUC Δ in $AIC = 7.6$, $X^2(1) = 9.60$, $p = .002$), but the model fit for MD did not reach significance (Δ in $AIC = 1.7$, $X^2(1) = .36$, $p = .548$). The results from the Likelihood Ratio Tests indicate that semantic knowledge trial-types did predict response duration and AUC scores, but did not predict MD scores. Table 2 reports the output from the full GLMMs.

Table 2

Generalized linear mixed model results for trial-type in the semantic knowledge contrast.

Model Type		Intercept			Primary Fixed Factor (Trial Type)		
	n	Estimate	SE	t	Estimate	SE	t
Duration ~ Trial-Type	2280	2153.81	60.30	35.72 ***	117.13	38.24	3.06 **
AUC ~ Trial-Type	2280	.37	.04	8.45 ***	.12	.04	3.10 **
MD ~ Trial-Type	2280	.23	.02	13.89 ***	.01	.01	.60

Note. SE = Standard Error. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 2 shows that the average response duration was 2154ms and the average AUC was .37 units^2 in the SK-proximal trials. Table 2 continues to show that response duration significantly increased on average 117ms and AUC significantly increased on average .12 units^2 in SK-distal trials. The significant increase shows that participants were slower and created bigger curves for the SK-distal trials. The performance differences between SK-proximal and SK-distal trials indicate that SK-distal trials placed higher demands on semantic control.

The semantic knowledge trial-types did not predict MD scores due to the insignificant Likelihood Ratio Test. However, Table 2 shows that the average MD scores for SK-proximal trials was .23 units and the mean difference between SK-proximal and SC-distal trials was .01 units.

3.2.2 Semantic Control Contrast

3.2.2.1 Accuracy. Figure 5a shows that SC-distal trials were more accurate compared to SC-proximal trials. The mean difference was 4.08%. Identical to the semantic knowledge contrast, the accuracy scores in the semantic control contrast were not normally distributed and a Wilcoxon Signed Rank Test was used to test for a statistical difference between SC-proximal and SC-distal trials. A significant difference was found for accuracy scores between SC-proximal and SC-distal (z = -4.24, $p < .001$, r = -.47). The analysis indicated that SC-distal trials were on average more accurate than SC-proximal trials and that the size of the effect was large.

3.2.2.2. Response Duration and Geometric Mouse-tracking Scores. Three GLMMs were used to assess the effect of trial-type for the semantic control contrast in terms of response duration and for geometric measures of mouse curvature (AUC and MD). Trial-type was entered as a fixed factor and participant was entered as a random factor.

Three Likelihood Ratio Tests were computed to compare the GLMMs to random-effects only models (GLMMs with no fixed factor). All three model comparisons were significant (Duration Δ in AIC = 170, $X^2(1)$ = 171.14, $p < .0001$; AUC Δ in AIC = 61.7, $X^2(1)$ = 63.68, $p < .0001$; MD Δ in AIC = 73.5, $X^2(1)$ = .75.49, $p < .0001$). The results from the Likelihood Ratio Tests indicate that semantic control trial-types did predict response duration, AUC scores and MD scores on the SPM. Table 3 reports the output from the GLMMs.

Table 3

Generalized linear mixed model results for trial-type in the semantic control contrast.

Model Type		Intercept			Primary Fixed Factor (Trial Type)		
	n	Estimate	SE	t	Estimate	SE	t
Duration ~ Trial-Type	2191	2077.57	67.01	31.00 ***	566.20	42.41	13.35 ***
AUC ~ Trial-Type	2191	.40	.05	8.13 ***	.36	.04	8.04 ***
MD ~ Trial-Type	2191	.22	.02	12.47 ***	.13	.02	8.77 ***

Note. SE = Standard Error. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 3 shows that, on average, all SPM variables significantly increased for SC-proximal trials compared to SC-distal trials. The average response duration was 2078ms, the average AUC was .40 units^2 and the average MD was .22 units in the SC-distal trials. The averages significantly increased to 2644ms, .76 units^2 and .35 units respectively in the SC-proximal trials. The increase shows that participants were slower, created bigger curves and deviated further in SC-proximal trials. The performance differences between SC-distal and SC-proximal trials indicate that SC-proximal trials placed higher demands on semantic control.

Now that GLMMs have been established with trial-type as the only fixed factor for both the semantic knowledge and semantic control contrasts, these acted as null models to understand the involvement of age, vocabulary and cognitive control predictors in the SPM. Age and vocabulary GLMMs will be discussed next.

3.3 Exploring Semantic Control in the SPM with Age and Vocabulary

The second hypothesis seeks to test whether age and vocabulary knowledge can separately influence a child's performance during the SPM. If a child is older or has a broader vocabulary, they may be able to select the most semantically related alternative quicker (lower duration) and more precisely (lower AUC and MD) due to better semantic control abilities. In terms of the GLMMs already reported, this third section is investigating whether the fixed effects of trial-type are still stable when accounting for participant age and vocabulary knowledge in the model. The second hypothesis will be examined using the response durations and geometric mouse-tracking scores (AUC and MD), as well as chronological age and BPVSIII raw scores.

The BPVSIII raw scores ranged from 84 to 161 (*M* = 126.1, *SD* = 17.55, *n* = 78) and were centred to then be added as a continuous variable into the GLMMs. Chronological age was added as a categorical variable; group one signifies 6- to 7-year-olds (*M* = 7 years 4 months, *SD* = 5.46 months, *n* = 29) and group two signifies 8- to 12-year-olds (*M* = 9 years 7.64 months, *SD* = 1 year .66 months, *n* = 50). Figure 6 shows the age group performance differences on the SPM task.

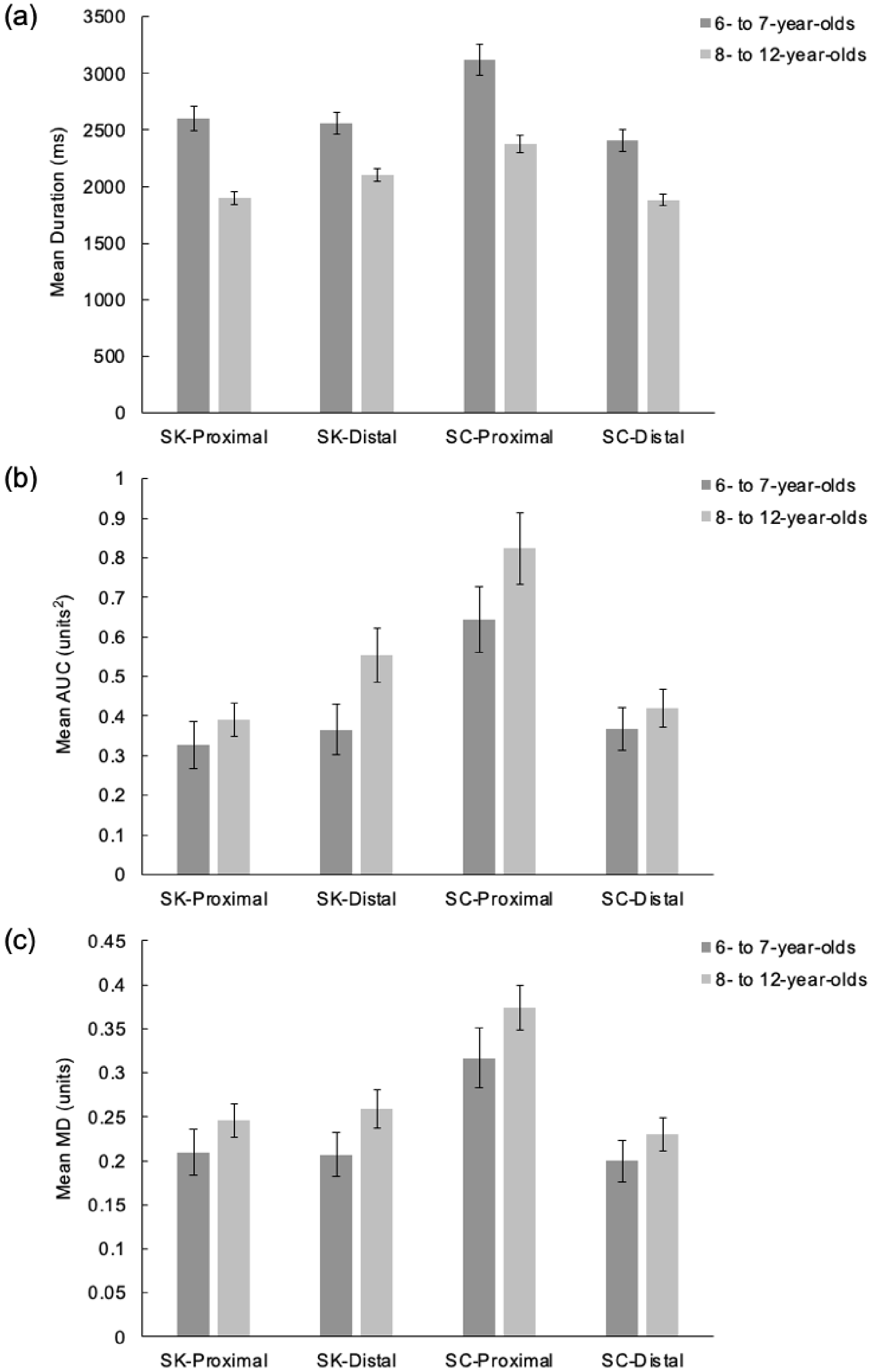

Figure 6. Mean response duration, AUC scores and MD scores for the SPM trial-types with standard error as error bars.

Figure 6a shows that younger children were on average half a second slower at making their alternative choice across all trial-types. Both age groups took the longest on SC-proximal trials, but older children's shortest response duration was for SK-proximal trials whereas younger children's shortest response time was SC-distal.

Figure 6b shows that older children created on average bigger curves compared to younger children on all trial-types. Both age groups additionally had the biggest curves for SC-proximal trials and the smallest curves for SK-proximal trials.

Figure 6c shows that older children on average had bigger deviations across all trial-types compared to younger children. Even though the older age group had larger average MD, both age groups had the same pattern for the largest and smallest average MD; SC-proximal trials created the biggest and SC-distal created the smallest. However, the second largest MD differed by age. Younger children's second biggest average MD was SK-proximal trials and older children's second biggest average MD was SK-distal.

Overall, older and younger children behaved similarly across all variables and were therefore pooled into one analysis and not into two separate analyses.

3.3.1 Semantic Knowledge Contrast

Six GLMMs were used to assess the effect of chronological age and vocabulary knowledge on trial-type performance for the semantic knowledge contrast in terms of response duration and geometric measures of mouse curvature (AUC and MD).

The first three GLMMs entered trial-type and age group as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and age GLMMs to the trial-type only GLMMs. The model comparison with duration as the dependent variable was significant (Δ in $AIC = 26$, $X^2(1) = 27.83$, $p < .0001$), but the AUC and MD models were non-significant (AUC Δ in $AIC = .5$, $X^2(1) = 2.54$, $p = .111$; MD Δ in $AIC = 0$, $X^2(1) = 2.01$, $p = .157$). The results from the Likelihood Ratio Tests indicate that chronological age did predict response duration in the semantic knowledge trial-types, but did not predict AUC or MD scores.

The second three GLMMs entered trial-type and centred BPVSIII raw score as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and vocabulary GLMMs to the trial-type only GLMMs. The model comparisons including duration as the dependent variable was significant (Δ in *AIC* = 10, $X^2(1) = 12.57$, $p < .001$), but the AUC and MD models were non-significant (AUC Δ in *AIC* = 1.9, $X^2(1) = .12$, $p = .732$; MD Δ in *AIC* = 1.8, $X^2(1) = .12$, $p = .73$). The results from the Likelihood Ratio Tests indicate that vocabulary knowledge did predict response duration in the semantic knowledge trial-types, but did not predict AUC or MD scores. Table 4 reports the output from the GLMMs with age and vocabulary.

Table 4

Generalized linear mixed model results for age and vocabulary regarding trial-type in the semantic knowledge contrast.

Model Type		Intercept			Primary Fixed Factor (Trial Type)			Secondary Fixed Factor		
	n	Estimate	SE	t	Estimate	SE	t	Estimate	SE	t
Duration ~ Trial-Type + Age	2280	3094.49	171.21	18.07 ***	116.78	38.24	3.05 **	-575.60	99.78	-5.77 ***
Duration ~ Trial-Type + BPVS	2267	2148.10	56.42	38.08 ***	117.20	38.28	3.06 **	-11.24	3.05	-3.69 ***
AUC ~ Trial-Type + Age	2280	.16	.14	1.14	.12	.04	3.11 **	.13	.08	1.61
AUC ~ Trial-Type + BPVS	2267	.37	.04	8.39 ***	.12	.04	3.06 **	- .001	.002	-.34
MD ~ Trial-Type + Age	2280	.16	.05	2.91 **	.01	.01	.61	.05	.03	1.43
MD ~ Trial-Type + BPVS	2267	.23	.02	13.75 ***	.01	.01	.58	- .001	.001	-.35

Note. SE = Standard Error. * p < .05, ** p < .01, *** p < .001

In terms of age, Table 4 shows that response duration still significantly increased on average 117ms and AUC significantly increased on average .12 units^2 for SK-distal trials compared to SK-proximal when controlling for age. However, the intercept in the GLMM for AUC was non-significant, which means that the deviation in AUC scores is not beyond chance and should be interpreted with caution. In addition to the stable trial-type differences, children aged 8 to 12 years have on average a significantly lower response duration of 576ms compared to children aged 6 to 7 years. The lower response duration means that older children were on average faster at picking the picture-targets compared to younger children in the semantic knowledge contrast. No age differences were found for AUC or MD scores.

A similar pattern can be shown in terms of vocabulary knowledge. Table 4 shows that response duration still increased on average 117ms and AUC increased on average 12 units^2 for SK-distal trials compared to SK-proximal when controlling for vocabulary knowledge. Table 4 continues to show that when children's BPVS raw score increased by one point, their response duration significantly decreased by 11ms. The lower response duration means that children with better vocabulary scores were on average faster at picking the picture-targets compared to children with lower vocabulary scores in the semantic knowledge contrast. No differences were found for AUC or MD scores when incorporating vocabulary knowledge.

3.3.2 Semantic Control Contrast

Six GLMMs were used to assess the effect of chronological age and vocabulary knowledge on trial-type performance for the semantic control contrast in terms of response duration and geometric measures of mouse curvature (AUC and MD).

The first three GLMMs entered trial-type and age group as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and age GLMMs to the trial-type only GLMMs. The model comparison with duration as the dependent variable was significant (Δ in AIC = 25, $X^2(1) = 27.43$, $p < .0001$), but the AUC and MD models were non-significant (AUC Δ in AIC = .2, $X^2(1) = 1.79$, $p = .181$; MD Δ in AIC = 0, $X^2(1) = 1.92$, $p = .166$). The results from the Likelihood Ratio Tests indicate that chronological age did predict response duration in the semantic control trial-types, but did not predict AUC or MD scores.

The second three GLMMs entered trial-type and centred BPVSIII raw score as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and vocabulary GLMMs to the trial-type only GLMMs. The model comparison with duration as the dependent variable was significant (Δ in *AIC* = 6, $X^2(1) = 8.50$, $p < .004$), but the AUC and MD models were non-significant (AUC Δ in *AIC* = 1.9, $X^2(1) = .07$, $p = .791$; MD Δ in *AIC* = 2, $X^2(1) = .0002$, $p = .988$). The results from the Likelihood Ratio Tests indicate that vocabulary knowledge did predict response duration in semantic control trial-types, but did not predict AUC or MD scores. Table 5 reports the output from the GLMMs with age and vocabulary.

Table 5

Generalized linear mixed model results for age and vocabulary regarding trial-type in the semantic control contrast.

Model Type		Intercept			Primary Fixed Factor (Trial Type)			Secondary Fixed Factor		
	n	Estimate	SE	t	Estimate	SE	t	Estimate	SE	t
Duration ~ Trial-Type + Age	2191	3118.26	190.42	16.38 ***	567.37	42.41	13.38 ***	-637.25	111.09	-5.74 ***
Duration ~ Trial-Type + BPVS	2178	2074.63	64.38	32.23 ***	566.33	42.39	13.36 ***	-10.48	3.50	-3.00 **
AUC ~ Trial-Type + Age	2191	.20	.16	1.28	.35	.04	8.03 ***	.12	.09	1.35
AUC ~ Trial-Type + BPVS	2178	.40	.05	8.01 ***	.36	.04	8.11 ***	.001	.003	.26
MD ~ Trial-Type + Age	2191	.14	.06	2.57 *	.13	.02	8.76 ***	.05	.03	1.40
MD ~ Trial-Type + BPVS	2178	.22	.02	12.35 ***	.13	.02	8.77 ***	.000	.001	.02

Note. SE = Standard Error. * p < .05, ** p < .01, *** p < .001

In terms of age, Table 5 shows that response duration still significantly increased on average 567ms, AUC still significantly increased .35 units^2 and MD still significantly increased .13 units for SC-proximal trials compared to SC-distal trials when controlling for age. However, the intercept in the GLMM for AUC was non-significant, which means that the deviation in AUC scores is not beyond chance and should be interpreted with caution. Table 4 continues to show that children aged 8 to 12 years have on average a significantly lower response duration of 637ms compared to children aged 6 to 7 years. The lower response duration means that older children were on average faster at picking the picture-targets compared to younger children in the semantic control contrast. No age differences for AUC or MD scores were revealed.

A similar pattern can be found in terms of vocabulary knowledge. Table 5 shows that response duration still significantly increased on average 566ms, AUC still significantly increased .36 units[2] and MD still significantly increased .13 units for SC-proximal trials compared to SC-distal trials when controlling for vocabulary knowledge. In addition to the stable trial-type differences, response duration significantly changed based on children's vocabulary ability. Table 5 shows that when children's BPVS raw score increased by one point, their response duration significantly decreased by 10ms. The lower response duration means that children with better vocabulary scores were on average faster at picking the picture-targets compared to children with lower vocabulary scores in the semantic control contrast. No differences were found for AUC or MD scores when incorporating vocabulary knowledge.

3.4 Exploring Semantic Control in the SPM with Cognitive Control

The third hypothesis seeks to test whether scores on the SPM were associated with performance on other measures of cognitive control. Maybe if a child has better cognitive control (higher Go/No-Go sensitivity and smaller WCST post switch perseveration error), they would be able to select the most semantically related alternative quicker (lower duration) and more precisely (lower AUC and MD). The third hypothesis will be examined using the response durations and geometric mouse-tracking scores, as well as Go/No-Go sensitivity and proportion of WCST post switch perseveration error. Table 6 shows the descriptive statistics for the cognitive control tasks.

Table 6

Mean performance and Standard Deviations for Go/No-Go (n = 79) and WCST (n = 78).

Go/No-Go			WCST Pre-Switch		WCST Post-Switch	
Go Accuracy	No-Go Accuracy	Sensitivity	Accuracy	Perseveration Error	Accuracy	Perseveration Error
99.2% (1.99)	90.7% (9.73)	8.14 (2.39)	79.81% (20.81)	17.88% (19.05)	60.00% (17.50)	23.92% (17.27)

Note.

Participants performed very well on the Go/No-Go task since average accuracy scores were high for both Go and No-Go trials. Due to the high accuracy rates, sensitivity was generally high as well (*Max* = 10.44, *Min* = 2.29).

Participants were less accurate on the WCST compared to the Go/No-Go. During the WCST, participants generally performed better in pre-switch trials compared to post-switch trials as seen by the perseveration errors in table 6.

3.4.1 Semantic Knowledge Contrast

Six GLMMs were used to assess the effect of Go/No-Go and WCST ability on trial-type performance for the semantic knowledge contrast in terms of response duration and geometric measures of mouse curvature (AUC and MD).

The first three GLMMs entered trial-type and Go/No-Go sensitivity as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and Go/No-Go GLMMs to the trial-type only GLMMs. The model comparison with duration as the dependent variable was significant (Δ in *AIC* = 12, $X^2(1)$ = 13.64 $p < .001$), but the AUC and MD models were non-significant (AUC Δ in *AIC* = 1.4, $X^2(1)$ = .61, p = .435; MD Δ in *AIC* = 1.8, $X^2(1)$ = .23, p = .629). The results from the Likelihood Ratio Tests indicate that Go/No-Go sensitivity did predict response duration, but did not predict AUC or MD scores in the semantic knowledge trial-types.

The second three GLMMs entered trial-type and centred WCST post-switch perseveration error as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and WCST GLMMs to the trial-type only GLMMs. None of the model comparisons were significant (duration Δ in *AIC* = 2, $X^2(1) = 3.72$, $p = .054$; AUC Δ in *AIC* = 1.7, $X^2(1) = .36$, $p = .546$; MD Δ in *AIC* = 1.5, $X^2(1) = .54$, $p = .462$). The results from the Likelihood Ratio Tests indicate that WCST perseveration error performance did not predict response duration, AUC scores or MD scores in the semantic knowledge trial-types. Table 7 reports the output from the GLMMs with Go/No-Go and WCST.

Table 7

Generalized linear mixed model results for Go/No-Go and WCST regarding trial-type in semantic knowledge contrast.

Model Type		Intercept			Primary Fixed Factor (Trial Type)			Secondary Fixed Factor		
	n	Estimate	SE	t	Estimate	SE	t	Estimate	SE	t
Duration ~ Trial-Type + Go/No-Go	2280	2851.58	189.54	15.05 ***	116.89	38.24	3.06 **	-85.6	22.21	-3.86 ***
Duration ~ Trial-Type + WCST	2252	2144.40	59.41	36.09 ***	124.01	38.40	3.23 **	6.361	3.26	1.95
AUC ~ Trial-Type + Go/No-Go	2280	.26	.14	1.85	.12	.04	3.10 **	.01	.02	.78
AUC ~ Trial-Type + WCST	2252	.37	.04	8.49 ***	.12	.04	3.04 **	.001	.002	.60
MD ~ Trial-Type + Go/No-Go	2280	.21	.06	3.70 ***	.008	.01	.60	.003	.007	.48
MD ~ Trial-Type + WCST	2252	.24	.02	13.96 ***	.006	.01	.49	.001	.001	.74

Note. SE = Standard Error. * p < .05, ** p < .01, *** p < .001

In terms of Go/No-Go, Table 7 shows that response duration still significantly increased on average 117ms and AUC significantly increased on average 12 units^2 for SK-distal trials compared to SK-proximal when controlling for Go/No-Go performance. However, the intercept in the GLMM for AUC was non-significant, which means that the deviation in AUC scores is not beyond chance and should be interpreted with caution. Table 7 continues to show that when children's Go/No-Go sensitivity score increased by one point, their response duration significantly decreased 86ms. The lower response duration means that children with better

Go/No-Go performance were on average faster at picking the target-picture compared to children with lower Go/No-Go performance in the semantic knowledge contrast. No performance differences in AUC or MD scores were established.

In terms of WCST, Table 7 shows that response duration still significantly increased on average 124ms and AUC significantly increased on average .12 units^2 for SK-distal trials compared to SK-proximal trials when controlling for WCST performance. No difference was found for MD scores. Additionally, WCST performance did not help predict any change in response duration, AUC scores or MD scores.

3.4.2 Semantic control contrast

Six GLMMs were used to assess the effect of Go/No-Go and WCST ability on trial-type performance for the semantic control contrast in terms of response duration and geometric measures of mouse curvature (AUC and MD).

The first three GLMMs entered trial-type and Go/No-Go sensitivity as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and Go/No-Go GLMMs to the trial-type only GLMMs. The model comparison with duration as the dependent variable was significant (Δ in *AIC* = 10, $X^2(1) = 12.82$, $p < .001$), but the AUC and MD models were non-significant (AUC Δ in *AIC* = 1.7, $X^2(1) = .34$, $p = .560$; MD Δ in *AIC* = 1.6, $X^2(1) = .33$, $p = .568$). The results from the Likelihood Ratio Tests indicate that Go/No-Go sensitivity did predict response duration, but did not predict AUC or MD scores in the semantic control trial-types.

The second three GLMMs entered trial-type and centred WCST post-switch perseveration error as fixed factors. Participant was entered as a random factor. Three Likelihood Ratio Tests were computed to compare the trial-type and WCST GLMMs to the trial-type only GLMMs. None of the model comparisons were significant (duration Δ in *AIC* = 0, $X^2(1) = 2.21$, $p = .138$; AUC Δ in *AIC* = 1.6, $X^2(1) = .04$, $p = .529$; MD Δ in *AIC* = 1.2,

$X^2(1) = .81$, $p = .368$). The results from the Likelihood Ratio Tests indicate that WCST perseveration error performance did not predict response duration, AUC scores or MD scores in the semantic control trial-types. Table 8 reports the output from the GLMMs with Go/No-Go and WCST.

Table 8

Generalized linear mixed model results for Go/No-Go and WCST regarding trial-type in semantic control contrast.

Model Type		Intercept			Primary Fixed Factor (Trial Type)			Secondary Fixed Factor		
	n	Estimate	SE	t	Estimate	SE	t	Estimate	SE	t
Duration ~ Trial-Type + Go/No-Go	2191	2832.09	211.70	13.38 ***	566.67	42.41	13.36 ***	-92.63	24.83	-3.73 ***
Duration ~ Trial-Type + WCST	2163	2068.66	65.85	31.42 ***	559.67	42.57	13.15 ***	5.41	3.62	1.50
AUC ~ Trial-Type + Go/No-Go	2191	.31	.16	1.94	.36	.04	8.04 ***	.01	.02	.58
AUC ~ Trial-Type + WCST	2163	.40	.05	8.09 ***	.36	.05	8.12 ***	.002	.003	.63
MD ~ Trial-Type + Go/No-Go	2191	.19	.06	3.26 **	.13	.02	8.76 ***	.004	.007	.57
MD ~ Trial-Type + WCST	2163	.22	.02	12.33 ***	.14	.02	8.90 ***	.001	.001	.90

Note. SE = Standard Error. * $p < .05$, ** $p < .01$, *** $p < .001$

In terms of Go/No-Go, Table 8 shows that response duration still significantly increased on average 567ms, AUC significantly increased on average .36 units^2 and MD significantly increased .13 units for SC-proximal trials compared to SC-distal trials when controlling for Go/No-Go performance. However, the intercept in the GLMM for AUC was non-significant, which means that the deviation in AUC scores is not beyond chance and should be interpreted with caution. Table 8 continues to show that when children's Go/No-Go sensitivity score increased by one point, their response duration significantly decreased 93ms. The lower response duration means that children with better Go/No-Go performance were on average faster at picking the picture-target compared to children with lower Go/No-Go performance in the semantic control contrast. No such difference was found for AUC or MD scores.

In terms of WCST, Table 8 shows that response duration still significantly increased on average 560ms, AUC significantly increased on average .36 units^2 and MD significantly increased on average .14 units for SC-proximal trials compared to SC-distal trials when controlling for WCST performance. However, WCST performance did not help predict performance differences in response duration, AUC scores or MD scores.

4. Discussion

The current project was a first attempt at measuring semantic control in typically-developing school-aged children using a computerised taxonomic two alternative forced choice semantic picture matching task (SPM) that tracked performance with mouse movements. The project aimed to determine whether school-aged children showed any evidence of engaging in semantic control processes through the within-trial manipulations of the SPM, whilst the SPM controlled for the semantic representations of the concepts. Furthermore, the project aimed to determine whether children's performance on the SPM was related to their chronological age, vocabulary, or cognitive control abilities. The findings are discussed below according to the hypotheses.

The first hypothesis predicted that school-aged children would engage in semantic control when completing the SPM due to the different within-trial control manipulations. More specifically, children were hypothesised to recruit more semantic control for the SK-distal trial-types compared to the SK-proximal trial-types due to the competition between two picture-alternatives with weak semantic associations to the picture-probes (semantic knowledge contrast), and children were hypothesised to recruit more semantic control for the SC-proximal trial-types compared to the SC-distal trial-types due to the competition between two picture-alternatives with strong semantic associations to the picture-probes (semantic control contrast). The increased recruitment of semantic control for the SK-distal and SC-proximal trial-types were predicted to result in less accuracy, slower response durations, bigger AUC scores and larger MD scores in mouse-tracking performance. In the semantic knowledge contrast, children were significantly less accurate, slower to select the correct response when they were accurate and produced bigger mouse-curves for the SK-distal trials compared to the SK-proximal trials. Contrary to the first hypothesis, no significant difference was observed for the MD scores in the semantic knowledge contrast. The effect size for accuracy was medium. In the semantic control contrast, children were significantly less accurate, slower to select the correct response when they were accurate, produced bigger mouse-curves scores and deviated farther from the picture-target for the SC-proximal trial-types compared to the SC-distal trial-types, confirming the hypothesis. The effect size for accuracy here was medium to large.

Overall, the findings supported the first hypothesis that the within-task control manipulations would engage semantic control in school-aged children when they had to select the most semantically related picture-alternative to a picture-probe. This is because the results from both contrasts suggest that children were able to select the most semantically related picture-target on the majority of trials, whilst modifying the speed and geometric mouse-tracking movements of their performance according to the degree of semantic control needed. The findings are consistent with evidence from previous studies changing the context of the task demands on a semantic picture-matching task, where even well-established concepts in semantic memory can elicit changes in performance (e.g., Bozeat et al., 2000). However, the current findings are novel for observing a difference with the use of mouse-tracking measures with school-aged children. Furthermore, the semantic control contrast was identified as placing more demands for semantic control compared to the semantic knowledge contrast due to the larger differences between the trial-types. Since the picture-probes were matched in both contrasts, the differences in performance between the contrasts suggested that the SPM was more than a reflection of word knowledge relating to the picture-probes because otherwise performance would have been similar in the contrasts.

An important strength of the SPM task was the use of geometric mouse-tracking scores to supplement the more typical measures of trial accuracy and response durations. Geometric mouse-tracking data can give important insights into how a decision is made through the unconscious movements of the computer mouse (Freeman & Ambady, 2010). A confident or direct cursor response (straight from picture-probe to picture-target) would reduce the AUC and MD scores, whereas a swerving or fluctuating cursor response would increase the AUC and MD scores. Although the geometric data mostly supported the first hypothesis, it was surprising that only the AUC scores differed from chance in the semantic knowledge contrast and not MD scores as well. The lack of a difference in average MD indicates that while there is a variation in size of the mouse-curve, the point that was furthest away from the most ideal trajectory did not significantly change. There could be a couple of reasons for why this happened. First, it could be that the type of mouse-curve used in the SK-proximal and SK-distal trial-types were different types of curves, which can create variations in AUC scores but not always variations in MD scores (see Appendix B). For example, a mouse-trajectory that swerves towards the picture-distractor for less time but to the same magnitude could produce a similar MD score compared to a mouse-trajectory that swerves towards the picture-distractor for more time but the same magnitude. In fact, AUC scores can mask the individual variations in curve strategies because of the way in which AUC is calculated (Hindy et al., 2009). For instance, a child who starts off in a straight line between the two picture-alternatives and makes a late decision to swerve to the picture-target will not have the same curve strategy as a child who starts off swerving towards to the picture-distractor and then decides to swerve to the picture-target mid-cursor movement, yet these different curves can still produce the same AUC scores. Second, it could be that the mouse-trajectories were on average similar across the SPM trial-types, but the differences in MD scores were more pronounced in the semantic control contrast. In fact, the results did show that there was a miniscule difference in MD for the semantic knowledge contrast (.01 units) and a relative bigger difference in MD for the semantic control contrast (.13). A closer analysis of the exact curves could provide further insight to the differences in mouse-trajectories, such as analysing the type of curves that are produced (Maldonado, Dunbar & Chelma, 2019). However, both possible explanations converge on a cautious interpretation that the average difference between the AUC scores for the semantic knowledge contrast is linked to the need for more semantic control in the SK-distal trials.

The findings for the first hypothesis are additionally consistent with Hoffman's findings using a word-matching task with older and younger adults. The trials in the word-matching task that placed greater demands on semantic control had lower accuracy scores and slower response durations (Hoffman, 2018). Interestingly, despite the design differences, there are some additional attributive consistencies between the findings from the SPM with children and the findings from the word-matching task with adults. Part of Hoffman's design aimed to differentiate two aspects of semantic control, semantic selection (selecting between competing alternatives) and controlled retrieval (retrieving a semantic meaning), and although the present study did not specifically distinguish these two semantic control processes, it is clear that the within-trial manipulations for the SPM contrasts could be construed measuring these two separate semantic control processes. Within the semantic knowledge contrast, the SK-distal trials required participants to choose between a distal and an unrelated picture-alternative in order to select a correct match (distal picture-alternative). Here, both picture-alternatives were concepts that likely do not automatically come to mind when seeing the picture-probe and it could therefore be argued that participants had to engage in controlled retrieval in order to identify the most-semantically related picture-alternative. With this theory in mind, it is likely that engaging in a controlled retrieval process for SK-distal trials generates the delay in the response durations and the bigger AUC scores compared to SK-proximal trials where the closer proximity of the picture-target would imply a reduced need, or no need at all, to engage in controlled retrieval. The contribution to the differences in AUC scores in particular indicates that the SK-distal trials not only required more processing speed or general ability based on response durations, but that participants were actively more distracted by the unrelated picture-alternatives and needed to resolve the distraction to make a correct choice. Within the semantic control contrast, the SC-proximal trials required participants to choose between a very proximal and a proximal picture-alternative in order to select a correct match (very proximal picture-alternative). Here, both picture-alternatives were concepts that likely do automatically come to mind when seeing the picture-probe and it could therefore be argued that participants had to engage in semantic selection in order to identify the most-semantically related picture-alternative. With this theory in mind, it is likely that engaging in a semantic selection process for SC-proximal trials generates the delay in the response durations, the bigger AUC scores and larger MD scores compared to SC-distal trials where the weaker semantic association of the picture-distractor would imply a reduced need, or no need at all, to engage in semantic selection. The contribution to the differences in AUC and MD scores again indicates that the SC-proximal trials not only required more processing speed or general ability based on response durations, but that participants were actively more distracted by the proximal picture-alternatives and needed to resolve the distraction to make a correct choice. Moreover, the separate semantic control theory suggests that children are more sensitive to semantic selection compared to controlled retrieval, which means that children have the same preference as the older adults in Hoffman's task. This further supports research that suggests that older adults and children share a similar level of cognition, but while children's cognition will progress with age, cognition in older adults will deteriorate (e.g., McCormack, Brown, Maylor, Darby & Green, 1999).

The second hypothesis predicted that variation in performance across the trial-types would be linked to the differences in participant chronological age and/or vocabulary knowledge. More specifically, if developmental factors account for performance differences in the SPM, then the differences between trial-types may disappear once age and vocabulary knowledge is controlled in the model. Moreover, younger children and children with less vocabulary knowledge were hypothesised to be slower, produce bigger AUC scores and produce larger MD scores compared to older children and children with more vocabulary knowledge.

When age-group, 6- to 7-year-olds and 8- to 12-year-olds, were included in the model fit, there was still a significant variation in response duration and AUC scores for the semantic knowledge contrast and still a significant variation in response duration, AUC scores and MD scores for the semantic control contrast. All children were still significantly slower and produced bigger mouse-curves in the SK-distal trial-types compared to the SK-proximal trial-types, and all children were still significantly slower, produced bigger mouse-curves and deviated farther from the picture-target in the SC-proximal trial-types compared to the SC-distal trial-types. Additionally, once the differences across trial-types were controlled, it was clear that age-group was a significant determinant of response duration. Older children were on average faster to respond with a correct response, regardless of any differences in trial-type. The exact reason for why older children benefitted from faster response durations was unclear, but three tentative explanations could be concluded. First, older children may have been faster at selecting the picture-target because they have more experience with semantic associations compared to their younger peers due to being taught with helpful lexical input from teachers for longer (Larraín, 2016) and are better at identifying semantic associations (Sauzéon et al., 2004). The added familiarity with semantic associations could translate into having the knowledge to make quicker inferences for semantic associations between two pictures. Second, it could be speculated that older children benefit from faster cognitive processing speeds and could therefore formulate a response faster. Past research has identified that age negatively correlates to response times due to processing speeds in development (Droit-Volet & Zélanti, 2013). It is possible that closer scrutiny of the time taken by the older children to start their mouse-movement (initiation time) would give further insight to cognitive processing speeds since the initiation time was removed from overall response time per trial. Lastly, the quicker response durations could reflect that older children were simply more motorically skilled at navigating their decision with a hand-held computer mouse. This third explanation is more probable due to the visual correlation between response time on the dot clicking game and chronological age (see Figure 4); older children were generally faster at using the computer mouse to make a simple computer mouse response. Older children's quickness on the dot clicking game could therefore have been translated to the SPM as well.

When vocabulary knowledge was included in the model fit, there was still a significant variation in response duration and AUC for the semantic knowledge contrast and still a significant variation in response duration, AUC scores and MD scores for the semantic control contrast. All children were still significantly slower and produced bigger mouse-curves in the SK-distal trial-types compared to the SK-proximal trial-types, and all children were still significantly slower, produced bigger mouse-curves and deviated farther from the picture-target in the SC-proximal trial-types compared to the SC-distal trial-types. Additionally, once the differences across trial-types were controlled, it was clear that vocabulary knowledge was a significant determinant of response duration. Children with better vocabularies according to the BPVSIII were on average faster to respond with a correct response, regardless of any differences in trial-type. The exact reason for why children with better vocabularies benefitted from faster response durations cannot be fully determined based solely on the data from the experiment, but two tentative explanations were formed. First, children with broader vocabularies are typically exposed to more written and spoken words (Schmitt, 2014), and the added vocabulary knowledge could influence how quickly children could associate a semantic relation in the SPM task because they could recognize the concepts quicker. However, since many steps had been taken to ensure that the stimuli used were familiar concepts to children, by both test-based age-of-acquisition data and exposure data, and children performed near to ceiling across all trial-types, this suggestion is a less likely explanation. Second, having more vocabulary breadth could translate into having broader semantic networks that are more closely connected and dense. The closer connectivity of the semantic network could therefore result in quicker processing speed when selecting the most semantically related picture-alternative based on the density of the semantic network.

Overall, the findings for the second hypothesis are consistent with Hoffman's findings that age and vocabulary breadth affect response times in a word-matching task with manipulations of semantic control. However, it is interesting that while age and vocabulary knowledge affected response duration, the effects did not extend to performance changes in the geometric mouse-tracking data. In fact, if one closely inspects that raw differences by age for AUC and MD scores (see Figure 6), older children are actually creating bigger curves and deviating more to the distractor compared to their younger peers. Even though a statistical analysis has not been completed to make a quantitative assessment of the difference, the descriptive differences are contradictory to what the second hypothesis predicted; that younger children will create bigger curves and deviate more. It can be postulated that the differences between older and younger children for AUC and MD scores are due to the older children being more sensitive to the semantic control demands of the SPM and that younger children do not recruit semantic control as efficiently as older children in order to select a picture-target. However, it can also be postulated that the choice of mouse-trajectories was different for older and younger children. For instance, older children may have tried to predict the picture-target before fully processing the stimuli and started their trajectories straight to the picture-distractor before swerving to the picture-target, or older children could manipulate the mouse-movements better because of increased motor mobility. Again, further analyses into the specific type of curves used can provide more complete answers.

The third hypothesis predicted that variation in performance across the trial-types could be linked to the differences in participant inhibitory control and/or cognitive flexibility. More specifically, if general cognitive control processes account for performance differences in the SPM, then the differences between trial-types may disappear once inhibitory control and cognitive control is controlled in the model. Moreover, children with worse performance on the Go/No-Go or the WCST were hypothesised to be slower, produce bigger AUCs and produce larger MDs compared to children who had better performance on the Go/No-Go or WCST.

When Go/No-Go sensitivity was included in the model fit, there was still a significant variation in response duration and AUC scores for the semantic knowledge contrast and still a significant variation in response duration, AUC scores and MD scores for the semantic control contrast. All children were still significantly slower and produced bigger mouse-curves in the SK-distal trial-types compared to the SK-proximal trial-types, and all children were still significantly slower, produced bigger mouse-curves and deviated farther from the picture-target in the SC-proximal trial-types compared to the SC-distal trial-types. Additionally, once the differences across trial-types were controlled, it was clear that Go/No-Go sensitivity was a significant determinant of response duration. Children with more sensitivity on the Go/No-Go were on average faster to select the picture-target, regardless of any differences in trial-type. Two suggestions to explain why higher sensitivity on the Go/No-Go resulted in faster response durations were composed. The first explanation is that children with better inhibitory control are able to navigate their semantic networks with minimal distractions and regulate the semantic association quicker compared to children with worse inhibitory control. Whilst this explanation would be more in-line with the CSC framework, the differences in response durations were similar in the semantic knowledge contrast (86ms) and semantic control contrast (93ms), which suggests that performance was not all that different even though the semantic associations were. In fact, the semantic associations that were more difficult to differentiate, as shown by accuracy scores (semantic associations in the SC-proximal trial-type), had a bigger, albeit tiny, reduction in response duration (7ms). The second suggestion explains the effect of Go/No-Go sensitivity on response durations as the general ability to inhibit a picture-distractor, regardless of its semantic association. This second suggestion is more likely due to the near non-existent difference between the semantic knowledge contrast and semantic control contrast (7ms), however, the contrasts were never statistically compared to each other and so a non-quantitative difference between the two is just a theory.

When WCST perseveration error was included in the model fit, there was still a significant variation in response duration and AUC scores for the semantic knowledge contrast and still a significant variation in response duration, AUC scores and MD scores for the semantic control contrast. All children were still significantly slower and produced bigger mouse-curves in the SK-distal trial-types compared to the SK-proximal trial-types, and all children were still significantly slower, produced bigger mouse-curves and deviated farther from the picture-target in the SC-proximal trial-types compared to the SC-distal trial-types. Additionally, once the differences across trial-types were controlled, it was clear that WCST perseveration error was not a significant determinant of any performance on the SPM, dismissing the predictions.

Overall, the findings for the third hypothesis were mostly inconsistent with past research. First, the impact that Go/No-Go sensitivity had for both weak (semantic knowledge contrast) and strong associations (semantic control contrast) was a new finding compared to the lack of an inhibitory control effect for strong associations in pre-schoolers (Thibaut et al., 2011). A reason for this could be due to this study adopting a combination score of both incorrect and correct responses on inhibitory control as opposed to just incorrect responses. The combination score allowed both rewards for accurate performance and penalties for inaccurate performance to be taken into account. Second, the lack of an effect for cognitive flexibility on semantic control is inconsistent with past research with adults (Hoffman, 2018). A reason for this could be that the SPM tapped into both semantic selection and controlled retrieval, whilst Hoffman's word-matching task found that cognitive flexibility, measured by the WCST, only affected semantic selection ability. The lack of a distinction between semantic selection and controlled retrieval in the SPM task could have masked the effect of cognitive flexibility in semantic control.

In summary, the study has demonstrated that semantic control processes are evident in school-aged children, but the three hypotheses were only partially met. Hypothesis one demonstrated that semantic control can be found to affect accuracy, response duration and geometric mouse-tracking measures when controlling for individual differences, with the exception that MD was not affected in the semantic knowledge contrast (i.e., for weakly associated semantic relationships). Hypothesis two demonstrated that older children and children with higher vocabulary scores had faster response durations, but age and vocabulary knowledge did not affect geometric mouse-tracking measures. Hypothesis three demonstrated that children with better inhibitory control had faster response durations, but inhibitory control did not affect geometric mouse-tracking measures, and cognitive flexibility did not impact response durations or geometric mouse-tracking scores. The impact of age, vocabulary knowledge and inhibitory control on response durations all suggest that there are developmental variations in semantic control, but that these variations could not be captured with geometric mouse-tracking scores to the same extent. Whilst the study uncovered some interesting findings, there are some strengths and weaknesses that should be addressed when conducting future research with a similar design and for any future publications with the current data.

First, the adoption of mouse-tracking as a way of measuring responses has not been a frequent approach with school-aged children and this experiment showed promising applicability of using geometric mouse-tracking measures with developing populations. The measurements of AUC and MD worked well, and these scores did not seem to be influenced by hand preference or by geographical location of the picture-target. Typically, mouse-tracking studies exclude left handers (Freeman & Ambady, 2010), but this experiment did not use a left-hand exclusion criterion as children will not have necessarily established their final dexterity. By using mouse-tracking, behavioural data, other than accuracy and response time, was recorded to show the magnitude of the distraction towards the picture-distractor. Whilst response durations may be influenced by motor ability and computer experience, the geometric mouse-tracking data are valid enough to demonstrate a predictable pattern of differences in the SPM task. The use of geometric mouse-tracking scores was therefore a strength of the project. However, from the current investigation alone, it is undetermined whether participants response durations are more susceptible to their speed of processing, and it is not clear whether there is a trade-off between faster responding and less direct mouse-trajectories (such as demonstrated by older children). Therefore, future investigations should consider more fine-grained analyses in regards to analysing the exact mouse-curves used by children when a mouse-tracking technique is adopted.

Second, the overall design of the SPM was strategically planned to measure semantic control through the competitiveness of semantic associations between pictured concepts. The underlying assumption was that the pictures directly activate fully enriched concepts that are registered through a semantic hub and are not necessarily from the activation of lexical representations where conceptual knowledge is stored. Care was taken to ensure that the picture stimuli was age-appropriate, semantically distinct and visually unique, which were clear strengths in the design. Despite these measures, there are various challenges for interpreting the SPM measures as a level of semantic control. One challenge is that the taxonomic picture-matching approach created a high overlap in the number of similar features between two pictures with high semantic similarity (e.g., Labrador and Dachshund) compared to competing picture pairs with lower semantic similarity (e.g., Labrador and Lion). Therefore, the SPM could be construed as a feature-matching task instead. This issue was unavoidable with the stimuli-mode, but in order to minimize any possible feature-matching in the participants, the children were instructed to match by relatedness and were taught what relatedness meant with exemplars which emphasised categories and shared semantic attributes of pictured concepts (e.g., SPM instruction: "The Ice Lolly is related to the Ice Cream because they are both <u>cold foods</u> from the <u>freezer</u>. (…) The Ice Lolly is also related to the grapes because they are both <u>foods</u>."). Another challenge was that even though it was assumed that lexical representations were not activated due to the picture stimuli, children could still have used phonological recoding to identify the lexical form of the picture label and compare the lexical recoded forms to each other rather than relying on the featural information of the pictures. Phonological recoding is the translation of a concept by combining the spelling and sound of a word and matching it to similarly sounding words stored in memory (Joseph, 2004). If phonological recoding is the most useful in the SPM, then the semantic control contrast that has a very proximal semantic association between the picture-probe and picture-target should generate general labels that quickly specify the picture-target (e.g., 'dog' for Labrador and Dachshund). In fact, observationally, when some participants were completing the SPM they would justify their picture selection by saying "they are related because that is an animal and that is an animal". However, the semantic control contrast was more inaccurate and yielded slower response durations compared to the phonologically more difficult semantic knowledge contrast, which indicates that phonological recoding is not taking place for the SPM task. Therefore, it can be expected that a phonological recoding technique is not taking place. The last challenge with the SPM was that the probes were recycled and assigned to two trial-types in the design. Each probe was either a SK-proximal and SC-distal trial-type or a SK-distal and SC-proximal trial-type. Due to these pairings, a bias from the picture-probe to picture-target semantic relation could potentially arise for every probe since the trial-types had either consistently more separated picture-alternatives (SK-proximal and SC-distal) or consistently less separated picture-alternatives (SK-distal and SC-proximal). One way to address the picture-probe bias in future work is to generate additional presentation orders where each probe is assigned to the opposite trial-types. The added counterbalancing could be achieved in a between-group design as well to minimise repetition of the same probe for a single participant.

Third, and finally, another strength of the project was the use of GLMMs for the statistical analysis. The GLMM approach allowed the analyses to investigate whether fixed factors could predict performance in the SPM as well as taking random factors into consideration at the same time. Yet, different approaches to the GLMMs can be made with the current data in order to understand the effects of the fixed factors better and to control for more random factors that were present. For example, instead of adding just one secondary fixed factor for each GLMM, a tertiary fixed factor could be added at the same time to investigate the shared and independent variance between secondary and tertiary fixed factors. It would be particularly informative to add age to the vocabulary knowledge and cognitive control GLMMs as an interaction since such models could untangle whether vocabulary knowledge and cognitive control is led by age or not. In fact, vocabulary knowledge (Farkas & Beron, 2004), inhibitory control (Cragg & Nation, 2008) and cognitive flexibility (Chelune & Baer, 1986; Dick, 2014) are all skills that are positively affected by age in development, and therefore fixed factor interactions with age would uncover whether the vocabulary knowledge and cognitive control measures in the study reflect what they were measuring (i.e., vocabulary knowledge and cognitive control) or whether they reflect age through a different measure. Similarly, adding age as a continuous variable rather than a dichotomous variable is another approach that can be adopted for the GLMMs. Research suggests that a continuous age variable in regression analyses allows for models to have greater predictive power and the models will in turn indicate the expected change in performance for the dependent variables per added year or month rather than the change in performance based on age-group (SAGE publications, 2019, p. 5). Such an approach would additionally be more useful to explain the developmental trajectory of semantic control and age since it would incorporate every participant's exact age and not round it up or down. A last alternative approach to the data analysis would be to add the 60 different picture-alternative pairings as a second random factor in the GLMMs. Adding the SPM stimuli as a random factor would regulate any performance bias that may have been produced in practice for certain picture-alternative pairings despite the rigorous stimuli design. This approach would help to minimise any contamination made to the data based on the stimuli of every trial and it would help to understand whether the performance differences in the trial-types still remain when controlling for the stimuli. If the performance differences in the trial-types do not remain in any GLMMs with the picture-alternative pairings as a random factor, then the statistical analyses could indicate that there were certain picture combinations that led to the performance differences and not the required degree of semantic control.

Whilst there were strengths with the design and the data analysis, the mentioned weaknesses are just some points to address in future research and in any future publications using this experiment's collected data. There are additionally further research directions that future research can take in order to gain a better understanding about semantic cognition in children. A first research direction would be to create a thematic version of the SPM task. A thematic approach would entail that picture-probes and picture-targets are matched with a thematic, or event-based, relation and does not rely on feature similarities the same way that a taxonomic approach does. For example, high thematic pairings would have less featural similarities (e.g., pony and saddle, fork and broccoli, newspaper and coffee) compared to high taxonomic parings (e.g., pony and horse, fork and knife, newspaper and magazine). A manipulation of thematic relations may in fact uncover different findings for children in terms of their developmental associations, since ongoing research with thematic relations suggests that thematic matching requires more cognitive flexibility to generate the best match (Zhang et al., 2021).

Another future extension to take is to incorporate trial-types that were omitted completely in the current project, such as trial-types with proximal and distal picture-alternatives or trial-types with very proximal and unrelated picture alternatives. These added trial-types would allow for a more complete picture of semantic control as the former would place very high demands on semantic control and the latter would place near to no demands on sematic control. It would especially be useful to know whether accuracy for the proximal versus distal trial-type would be above chance in children as this could signify how advanced semantic control is.

A last future direction would be to adopt a semantic word-matching task to allow for a more direct comparison to Hoffman's adult study (2018). A taxonomic semantic word-matching task could be a good comparison for a taxonomic picture matching task to determine how well children are able to use their semantic representations since there are no cues for any feature resemblance directly from words (although phonetic trial-stimuli should be avoided due to the issues with phonological recoding).

In conclusion, the current study contributes important findings to the sparse literature on semantic cognition in children. When controlling for the requirement of semantic representations in the design of the SPM and controlling for extraneous developmental factors in the analysis, the manipulations of semantic control still led to robust patterns of findings in accuracy scores, response durations and geometric mouse-tracking scores. The results from the two contrasts imply that semantic control processes mediate children's picture-matching ability for both weak and strong associations, confirming that the CSC framework is a useful model for understanding semantic cognition in children. On top of the semantic control finding, chronological age, vocabulary knowledge and inhibitory control did contribute to differences in response durations for the SPM, but cognitive flexibility did not. Further inquiries into the development of semantic cognition should generate more insights into the role of cognitive control in conceptual knowledge. Such research can help educators and researchers develop learning strategies that can help children maximize their conceptual knowledge by applying it in relevant situational contexts, such as in the classroom.

References

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388-407. doi: 10.3758/s13428-019-01237-x

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R., & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease And Treatment*, 449. doi: 10.2147/ndt.s39776

Arias-Trejo, N., & Plunkett, K. (2009). Lexical-Semantic Priming Effects during Infancy. *Philosophical Transactions: Biological Sciences, 364*(1536), 3633-3647. Retrieved February 19, 2021, from http://www.jstor.org/stable/40538180

Badre, D., Hoffman, J., Cooney, J., & D'Esposito, M. (2009). Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature Neuroscience*, *12*(4), 515-522. doi: 10.1038/nn.2277

Badre, D., Poldrack, R., Paré-Blagoev, E., Insler, R., & Wagner, A. (2005). Dissociable Controlled Retrieval and Generalized Selection Mechanisms in Ventrolateral Prefrontal Cortex. *Neuron*, *47*(6), 907-918. doi: 10.1016/j.neuron.2005.07.023

Balota, D. A., & Coane, J. H. (2008). Semantic Memory. In J. H. Byrne (Ed.). *Learning and Memory: A Comprehensive Reference* (pp. 511-534), Academic Press.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi:10.18637/jss.v067.i01

Berg, E.A. (1948). Journal of Experimental Psychology, 38, 404-411. A simple objective technique for measuring flexibility in thinking. *Journal of Experimental Psychology, 39,* 15-22.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings Of The National Academy Of Sciences*, *109*(9), 3253-3258. doi: 10.1073/pnas.1113380109

Bishop, D., Ross, V., Daniels, M., & Bright, P. (1996). The measurement of hand preference: A validation study comparing three groups of right-handers. *British Journal Of Psychology*, *87*(2), 269-285. doi: 10.1111/j.2044-8295.1996.tb02590.x

Bozeat, S., Lambon Ralph, M., Patterson, K., Garrard, P., & Hodges, J. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*(9), 1207-1215. doi: 10.1016/s0028-3932(00)00034-8

Bozeat, S., Lambon Ralph, M., Graham, K., Patterson, K., Wilkin, H., & Rowland, J. et al. (2003). A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cognitive Neuropsychology*, *20*(1), 27-47. doi: 10.1080/02643290244000176

Brysbaert, M., & Biemiller, A. (2016). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*(4), 1520-1523. doi: 10.3758/s13428-016-0811-4

Chapman, C., Hasan, O., Schulz, P., & Martin, R. (2020). Evaluating the distinction between semantic knowledge and semantic access: Evidence from semantic dementia and comprehension-impaired stroke aphasia. *Psychonomic Bulletin & Review*, *27*(4), 607-639. doi: 10.3758/s13423-019-01706-6

Chelune, G., & Baer, R. (1986). Developmental norms for the wisconsin card sorting test. *Journal Of Clinical and Experimental Neuropsychology*, *8*(3), 219-228. doi: 10.1080/01688638608401314

Chen, Q., Garcea, F., & Mahon, B. (2015). The Representation of Object-Directed Action and Function Knowledge in the Human Brain. *Cerebral Cortex*, *26*(4), 1609-1618. doi: 10.1093/cercor/bhu328

Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. A. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: Evidence from the non-verbal domain. *Brain*, *132*(9), 2593-2608. https://doi.org/10.1093/brain/awp146

Cragg, L., & Nation, K. (2008). Go or no-go? Developmental improvements in the efficiency of response inhibition in mid-childhood. *Developmental Science*, *11*(6), 819-827. doi: 10.1111/j.1467-7687.2008.00730.x

De Deyne, S., Navarro, D., Perfors, A., Brysbaert, M., & Storms, G. (2018). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987-1006. doi: 10.3758/s13428-018-1115-7

Delacourte, A. (2009). Alzheimer Europe - Dementia - Other forms of dementia - Neurodegenerative diseases - Fronto-Temporal Degeneration - Semantic Dementia (SD). Retrieved 22 February 2021, from https://www.alzheimer-europe.org/Dementia/Other-forms-of-dementia/Neurodegenerative-diseases/Fronto-Temporal-Degeneration/Semantic-Dementia-SD

Dick, A. (2014). The development of cognitive flexibility beyond the preschool period: An investigation using a modified Flexible Item Selection Task. *Journal Of Experimental Child Psychology*, *125*, 13-34. doi: 10.1016/j.jecp.2014.01.021

Doebel, S. (2020). Rethinking Executive Function and Its Development. *Perspectives On Psychological Science*, *15*(4), 942-956. doi: 10.1177/1745691620904771

Doebel, S., & Zelazo, P. (2015). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, *38*, 241-268. doi: 10.1016/j.dr.2015.09.001

Droit-Volet, S., & Zélanti, P. (2013). Development of Time Sensitivity and Information Processing Speed. *Plos ONE*, *8*(8), e71424. doi: 10.1371/journal.pone.0071424

Dunn, L. M., & Dunn, D. M. (2009). *The British picture vocabulary scale*. GL Assessment Limited.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: differences by class and race. *Social Science Research*, *33*(3), 464-497. doi: 10.1016/j.ssresearch.2003.08.001

Freeman, J., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226-241. doi: 10.3758/brm.42.1.226

Foltz, P. W. (2001). Semantic Processing: Statistical Approaches. In N. J. Smelser & P.B. Baltes (Eds.). *International Encyclopedia of the Social & Behavioural Sciences* (pp. 13873-13878). Pergamon.

Gardner, H., Lambon Ralph, M., Dodds, N., Jones, T., Ehsan, S., & Jefferies, E. (2012). The Differential Contributions of pFC and Temporo-parietal Cortex to Multimodal Semantic Control: Exploring Refractory Effects in Semantic Aphasia. *Journal Of Cognitive Neuroscience*, *24*(4), 778-793. doi: 10.1162/jocn_a_00184

Hindy, N., Hamilton, R., Houghtling, A., Coslett, H., & Thompson-Schill, S. (2009). Computer-Mouse Tracking Reveals TMS Disruptions of Prefrontal Function During Semantic Retrieval. *Journal Of Neurophysiology*, *102*(6), 3405-3413. doi: 10.1152/jn.00516.2009

Hoffman, P. (2018). An individual differences approach to semantic cognition: Divergent effects of age on representation, retrieval and selection. *Scientific Reports*, *8*(1). doi: 10.1038/s41598-018-26569-0

Hoffman, P., McClelland, J., & Lambon Ralph, M. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293-328. doi: 10.1037/rev0000094

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., & Brugge, J. F. (2000). Auditory cortex on the human posterior superior temporal gyrus. *Journal of Comparative Neurology*, *416*(1), 79-92.

Jefferies, E. (2013). The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. *Cortex*, *49*(3), 611-625. doi: 10.1016/j.cortex.2012.10.008

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*(8), 2132-2147. https://doi.org/10.1093/brain/awl153

Johnson, E. B., Rees, E. M., Labuschagne, I., Durr, A., Leavitt, B. R., Roos, R. A., Reilmann, R., Johnson, H., Hobbs, N. Z., Langbehn, D. R., Stout, J. C., Tabrizi, S. J., & Scahill, R. I. (2015). The impact of occipital lobe cortical thickness on cognitive task performance: An investigation in Huntington's Disease. *Neuropsychologia*, *79*, 138-146.

Joseph, L. (2004). Reading Interventions. *Encyclopedia Of Applied Psychology*, 223-230. doi: 10.1016/b0-12-657410-3/00786-8

Laham, D., & Steinhart, D. (1998). LSA @ CU Boulder. Retrieved 22 February 2021, from http://lsa.colorado.edu

Lambon Ralph, M., Jefferies, E., Patterson, K., & Rogers, T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42-55. doi: 10.1038/nrn.2016.150

Landrigan, J., & Mirman, D. (2016). Taxonomic and Thematic Relatedness Ratings for 659 Word Pairs. *Journal Of Open Psychology Data*, *4*(1), 2. doi: 10.5334/jopd.24

Larraín, A. (2016). Argumentation and concept development: the role of imagination. *European Journal Of Psychology Of Education*, *32*(4), 521-536. doi: 10.1007/s10212-016-0316-7

Larraín, A., & Haye, A. (2014). A dialogical conception of concepts. *Theory & Psychology*, *24*(4), 459-478. doi: 10.1177/0959354314538546

Let's Go Learn. (2020). Retrieved 22 February 2021, from https://frontend.letsgolearn.com/practice/bubbles

Lovejoy, S. (2003). Printed Word Database for 5-9 Year Olds - Home. Retrieved 22 February 2021, from https://www1.essex.ac.uk/psychology/cpwd/

Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments & Computers, 36*(3), 421–431. https://doi.org/10.3758/BF03195590

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085-1101. doi: 10.3758/s13428-018-01194-x

McCormack, T., Brown, G., Maylor, E., Darby, R., & Green, D. (1999). Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology*, *35*(4), 1143-1155. doi: 10.1037/0012-1649.35.4.1143

Mirman, D., Landrigan, J., & Britt, A. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499-520. doi: 10.1037/bul0000092

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49-100. doi: 10.1006/cogp.1999.0734

Mummery, C., Patterson, K., Price, C., Ashburner, J., Frackowiak, R., & Hodges, J. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals Of Neurology*, *47*(1), 36-45. doi: 10.1002/1531-8249(200001)47:1<36::aid-ana8>3.0.co;2-l

Online Image Resizer. (2015). Retrieved 22 February 2021, from https://resizeimage.net/

Patterson, K., Nestor, P., & Rogers, T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976-987. doi: 10.1038/nrn2277

Pozueta, A., Lage, C., García-Martínez, M., Kazimierczak, M., Bravo, M., López-García, S., Fernández-Rodríquez, A., Riancho, J., González-Suárez, A., Vázquez-Higuera, J., L., Cano-Abascal, Á., Martínez-Dubarbie, F., Arcocha-Torres, M., Jiménez-Bonilla, J., Banzo, I., Rodríguez-Rodríguez, E., & Sánchez-Juan, P. (2020). A snake that bites its own tail. Acquisition and loss of concepts in children and semantic dementia patients through the analysis of drawings. *Cortex*, *128*, 162-173. doi: 10.1016/j.cortex.2020.03.007

PsychToolKit. (2021). Wisconsin Card Sorting Inspired Task (WCST). Retrieved 23 February 2021, from https://www.psytoolkit.org/experiment-library/wcst.html

R Core Team. (2020). A Language and environment for statistical computing. *R Foundation for statistical Computing*. Vienna, Austria, retrieved from https://www.R-project.org/

Rogers, T., & McClelland, J. (2006). *Semantic cognition*. Cambridge, Mass.: MIT.

Rogers, T., Patterson, K., Jefferies, E., & Lambon Ralph, M. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220-239. doi: 10.1016/j.neuropsychologia.2015.04.015

Romine C., Lee, D., Wolfe, M., Homack, S., George, C., & Riccio, C. (2004). Wisconsin Card Sorting Test with children: a meta-analytic study of sensitivity and specificity. *Archives Of Clinical Neuropsychology*, *19*(8), 1027-1041. doi: 10.1016/j.acn.2003.12.009

SAGE Publications. (2019). Learn About Analysing Age in Survey Data in SPSS with Data from the European Social Survey (2016). *SAGE Research Methods Datasets*, *2*, 1-21. Retrieved from http://methods.sagepub.com/base/download/DatasetStudentGuide/age-in-ess-2016-spss

Sánchez-Benavides, G., Gómez-Ansón, B., Quintana, M., Vives, Y., Manero, R. M., Sainz, A., Blesa, R., Molinuevo, J.L., & Peña-Casanova, J. (2010). Problem-solving abilities and frontal lobe cortical thickness in healthy aging and mild cognitive impairment. *Journal of the International Neuropsychological Society*, *16*(5), 836-845.

Sauzéon, H., Lestage, P., Raboutet, C., N'Kaoua, B., & Claverie, B. (2004). Verbal fluency output in children aged 7–16 as a function of the production criterion: Qualitative analysis of clustering, switching processes, and semantic network exploitation. *Brain And Language*, *89*(1), 192-202. doi: 10.1016/s0093-934x(03)00367-5

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, *64*(4), 913-951. doi: 10.1111/lang.12077

Snyder, H., & Munakata, Y. (2010). Becoming self-directed: Abstract representations support endogenous flexibility in children. *Cognition*, *116*(2), 155-167. doi: 10.1016/j.cognition.2010.04.007

Solomon, M., Ozonoff, S. J., Cummings, N., & Carter, C. S. (2008). Cognitive control in autism spectrum disorders. *International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience*, *26*(2), 239–247. https://doi.org/10.1016/j.ijdevneu.2007.11.001

Teige, C. (2017). *The neural basis of automatic and controlled semantic cognition: Spatiotemporal dynamics revealed by magnetoencephalography and brain stimulation* (Ph.D). The University of York.

Teige, C., Cornelissen, P., Mollo, G., Gonzalez Alam, T., McCarty, K., Smallwood, J., & Jefferies, E. (2019). Dissociations in semantic cognition: Oscillatory evidence for opposing effects of semantic control and type of semantic relation in anterior and posterior temporal cortex. *Cortex*, *120*, 308-325. doi: 10.1016/j.cortex.2019.07.002

Thibaut, J. P., French, R. M., Vezneva, M., Gérard, Y., & Glady, Y. (2011). Semantic analogies by young children: testing the role of inhibition. *European perspectives on cognitive science*, 136-140.

Thompson, H. E., Almaghyuli, A., Noonan, K. A., Barak, O., Lambon Ralph, M. A., & Jefferies, E. (2018). The contribution of executive control to semantic cognition: Convergent evidence from semantic aphasia and executive dysfunction. *Journal of neuropsychology*, *12*(2), 312-340.

Thompson-Schill, S., Ramscar, M., & Chrysikou, E. (2009). Cognition Without Control: When a Little Frontal Lobe Goes a Long Way. *Current Directions In Psychological Science*, *18*(5), 259-263. doi: 10.1111/j.1467-8721.2009.01648.x

Van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal Of Experimental Psychology*, *67*(6), 1176-1190. doi: 10.1080/17470218.2013.850521

Wainwright, P. E., Leatherdale, S. T., & Dubin, J. A. (2007). Advantages of mixed effects models over traditional ANOVA models in developmental studies: A worked example in a mouse model of fetal alcohol syndrome. *Developmental Psychobiology*, *49*(7), 664-674. https://doi.org/10.1002/dev.20245

Wiebe, S., Sheffield, T., & Espy, K. (2012). Separating the Fish From the Sharks: A Longitudinal Study of Preschool Response Inhibition. *Child Development*, *83*(4), 1245-1261. doi: 10.1111/j.1467-8624.2012.01765.x

Zhang, M., Varga, D., Wang, X., Krieger-Redwood, K., Gouws, A., Smallwood, J., & Jefferies, E. (2021). Knowing what you need to know in advance: The neural processes underpinning flexible semantic retrieval of thematic and taxonomic relations. *Neuroimage*, *224*, 117405. doi: 10.1016/j.neuroimage.2020.117405

Zortea, M., Menegola, B., Villavicencio, A., & Fumagalli de Salles, J. (2014). Graph analysis of semantic word association among children, adults, and the elderly. *Psicologia: Reflexão e Crítica*, *27*(1), 90-99. https://doi.org/10.1590/S0102-79722014000100011

Appendix A: Full list of SPM stimuli.

Probe	Very Proximal	Proximal	Distal	Unrelated
Apple	Pear	Broccoli	Egg	Tape
Bacon	Ham	Cheese	Banana	Balloons
Bar of Soap	Soap Dispenser	Washing up Liquid	Detergent tabs	Rabbit
Bee	Fly	Snail	Small Bird	Pan
Big Dog (Labrador)	Small Dog (Dachshund)	Fox	Lion	Table
Budgie	Parrot	Turkey	Penguin	Calculator
Carrot	Pepper	Pineapple	Cake	Cap
Chair	Sofa Chair	Bed	Bookcase	Sunglasses
Chocolate Bar	Chocolate Egg	Sweets	Biscuit	Book
Clock	Watch	Thermometer	Painting	Zebra
Computer	Laptop	Robot	Television	Tshirt
Coniferous Tree	Deciduous Tree	Bush	Flower	Umbrella
Cow	Scottish Highland Cow	Sheep	Donkey	Suitcase
Football	Basketball	Puck	Racket	Ring
Frog	Toad	Fish	Hedgehog	Money
Guitar	Violin	Keyboard	Flute	Shoes
Headphones	Earphones	Hairband	Hat2	Crayon
Helmet	Motorbike Helmet	Hat	Gloves	Crocodile
Manual Toothbrush	Electric Toothbrush	Hairbrush	Paintbrush	Newspaper
Motorbike	Moped	Bicycle	Tractor	Cat
Pencil	Pen	Eraser	Scissors	Rose
Plate	Cup	Glass	Pot	Goat
Pony	Horse	Pig	Deer	Jacket
Rucksack	Satchel	Plastic Bag	Pencil Case	Plant
Small Car (Suzuki)	Big Car (BMW)	Truck	Canoe	Lamp
Squirrel	Chipmunk	Rat	Bear	Golf Club
Swing	Tyre Swing	Slide	Bench	Pillow
Telephone	Mobile Phone	Camera	Radio	Door
Train	Tram	Ferry	Aeroplane	Vase
Trainers	Boots	Socks	Scarf	Ruler

Appendix B: Understanding AUC and MD as Geometric Mouse-Tracking Measures.

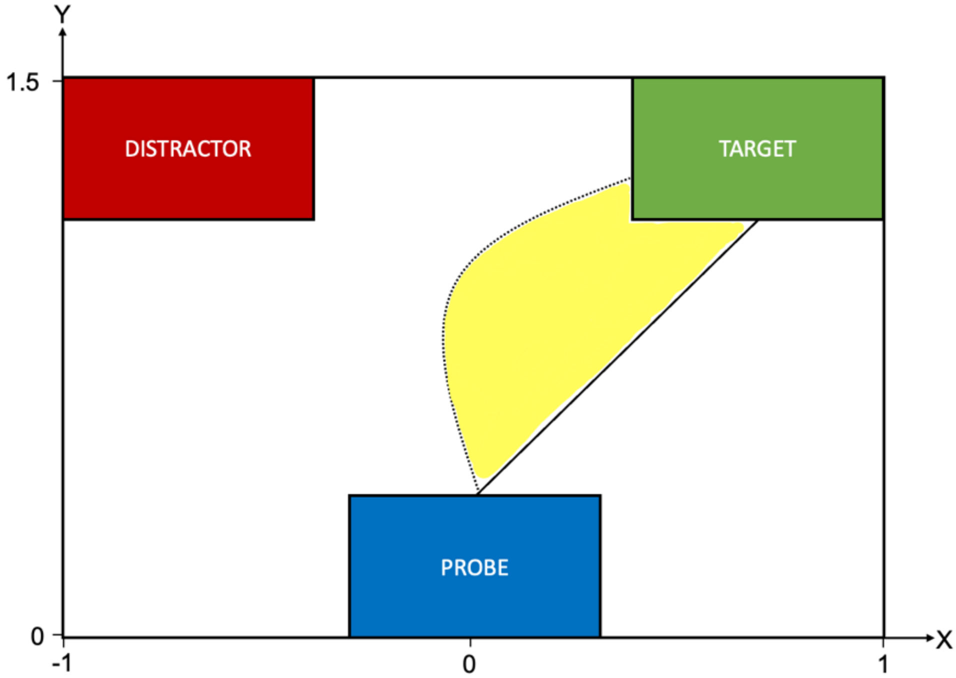

Figure showing that AUC is measured as the area under the curve (yellow) compared to the most ideal trajectory.

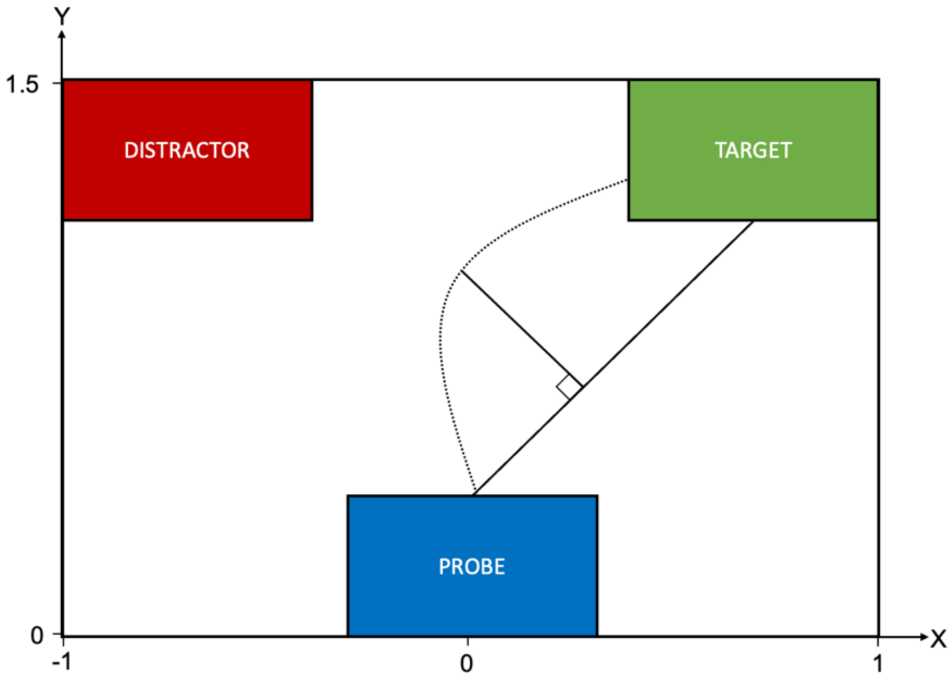

Figure showing that MD is measured as the point furthest away from the most ideal trajectory in a perpendicular direction.

Appendix C: Understanding Different Mouse-Curves.

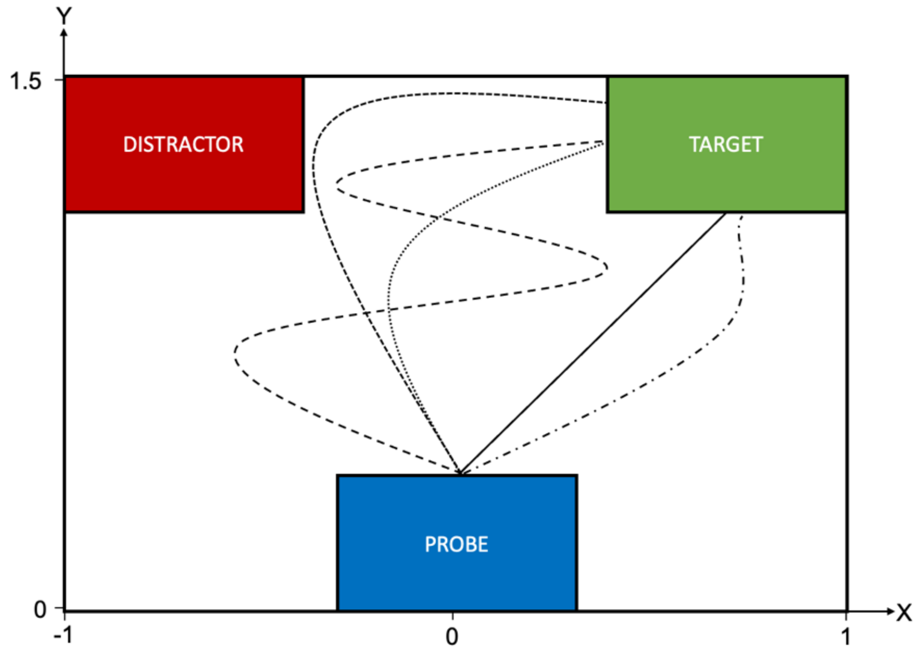

Figure showing different types of mouse trajectories that can be measured.

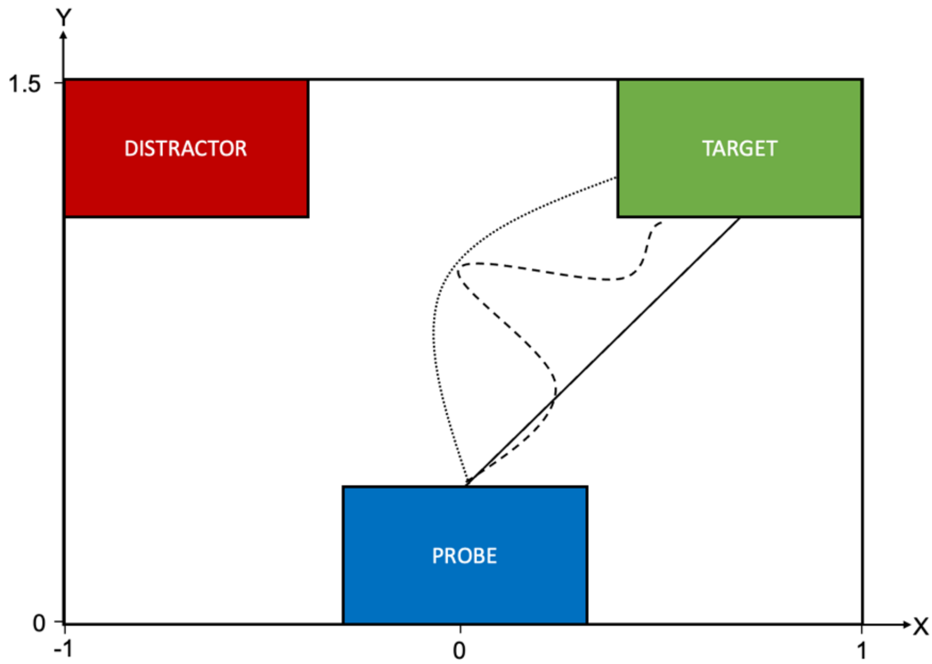

Figure showing how some trajectories that are different in AUC can still have the same MD.